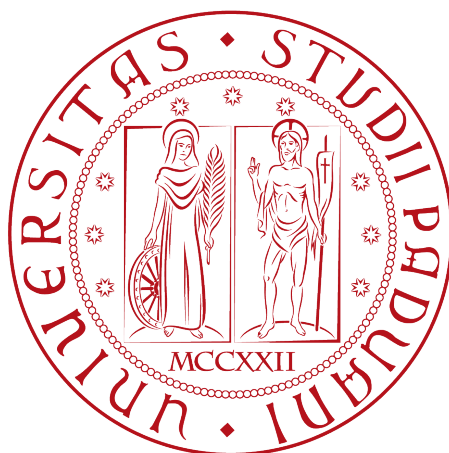


Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI LAUREA IN INFORMATICA



Analisi e gestione dei dati per il servizio di
raccolta rifiuti porta a porta presso Contarina
SpA e sviluppo di un modello predittivo

Tesi di laurea

Relatore

Prof. Luigi De Giovanni

Laureando

Giulio Zanatta
Matricola 1163163

ANNO ACCADEMICO 2022-2023

Giulio Zanatta: *Analisi e gestione dei dati per il servizio di raccolta rifiuti porta a porta presso Contarina SpA e sviluppo di un modello predittivo*, Tesi di laurea, ©
Febbraio 2023.

Dedicato alla mia famiglia, che non ha mai smesso di credere in me.

Sommario

La tesi espone il lavoro svolto durante l'attività di *stage* del laureando Zanatta Giulio presso l'azienda *Contarina SpA*, con una durata di 300 ore.

L'attività di *stage* prevede l'analisi e la comprensione dell'attuale sistema del servizio di raccolta rifiuti Porta a Porta (PaP) erogato da Contarina SpA, con lo scopo di effettuare previsioni future su specifici parametri di servizio tramite un modello predittivo. L'obiettivo principale è quello di fornire un sistema di stima dei parametri equivalente a quello esistente dal punto di vista operativo, ma più dinamico, volto al miglioramento del servizio e allo sviluppo di una proposta di riorganizzazione dello stesso.

Nei capitoli della tesi verrà introdotto e compreso il servizio erogato da *Contarina SpA*, attraverso un'analisi dei dati preliminari relativi alla produzione di rifiuti e la loro gestione. Successivamente, verranno presentati gli strumenti a disposizione dell'azienda per effettuare il servizio, seguiti da un approfondimento del piano in atto fino al giorno d'oggi per effettuare la raccolta.

Vi saranno dei capitoli relativi alla raccolta dei dati e relativa pulizia, seguiti dalla descrizione delle attività per l'analisi più approfondita che permette di individuare le variabili governabili del sistema PaP.

Infine, verrà sviluppata una proposta di riorganizzazione del servizio di raccolta attuale in base alle analisi effettuate e verrà sviluppato un modello predittivo.

Ringraziamenti

Innanzitutto, vorrei esprimere la mia gratitudine al Prof. Luigi De Giovanni, relatore della mia tesi, per l'aiuto e il sostegno fornitomi durante la stesura del lavoro.

Desidero ringraziare, soprattutto, con affetto i miei genitori e mia sorella per il sostegno, il grande aiuto e per aver sempre creduto in me in ogni momento durante gli anni di studio, perchè la forza del loro amore è stata imparagonabile.

Ho desiderio di ringraziare poi i miei veri amici per la vicinanza ricevuta durante questi anni, nonostante le difficoltà.

Ringrazio, inoltre, l'Ing. Stefano Valotto, tutor esterno, per l'aiuto nella stesura della mia tesi e per la sua disponibilità a fornire spiegazioni e per avermi fatto scoprire una realtà importante da diversi punti di vista.

Padova, Febbraio 2023

Giulio Zanatta

Indice

1	Introduzione	1
1.1	L'azienda	1
1.2	Proposta di stage	2
1.3	Descrizione dello stage	2
1.3.1	Obiettivi	3
1.3.2	Pianificazione del lavoro	3
2	Comprensione dello stato attuale	7
2.1	Produzione di rifiuti	7
2.2	Classificazione del territorio servito	8
2.3	Volumetrie	9
2.4	Il parco mezzi	9
2.4.1	Vaschette	9
2.4.2	Semirimorchi	10
2.5	Strumenti utilizzati	10
2.5.1	LeO (Localizzazione e Organizzazione)	11
2.5.2	ECOS	12
2.5.3	DBW	12
2.6	Pianificazione ed esecuzione del servizio PaP	12
2.6.1	Pianificazione	12
2.6.2	Esecuzione	14
2.6.3	Partecipazione ai turni di raccolta	15
2.7	Rendicontazione e raccolta dati del servizio PAP	17
3	Strategia di analisi e raccolta dati	21
3.1	Definizione del set di dati	21
3.2	Raccolta dei dati	22
3.3	Definizione delle basi dati	27
4	Analisi critica dei dati	31
4.1	Studio preliminare dei dati	31
4.2	Metodologia	31
4.3	Analisi Critica	32
4.3.1	Serie storiche	32
4.4	Approccio tradizionale all'analisi	32
4.5	Approccio moderno all'analisi	34
4.6	Analisi dati 12 mesi	35
4.6.1	Raccolta dati e definizione dataset	35

4.6.2	Analisi preliminare	36
4.6.3	Pulizia dei dataset	38
4.6.4	Analisi	42
4.7	Analisi dati 5 anni	46
4.7.1	Pulizia dei dati	46
4.7.2	Creazione dei dataset	47
4.7.3	Analisi	48
4.8	Modello predittivo	58
5	Riorganizzazione Operativa del servizio PAP	65
5.1	Algoritmo attuale	65
6	Conclusioni	69
6.1	Risultati ottenuti	69
6.2	Raggiungimento degli obiettivi	69
6.3	Consuntivo delle tempistiche	70
6.4	Conclusioni	71
	Riferimenti bibliografici	75

Elenco delle figure

1.1	Percentuale di raccolta differenziata, dati Contarina, [22], [6]	2
1.2	Algoritmo utilizzato per il calcolo delle risorse in uso tutt'oggi da Contarina SpA.	3
1.3	Diagramma di Gantt relativo alla pianificazione dell'attività di stage.	3
2.1	Mappa del territorio servito [20]	8
2.2	Suddivisione del territorio in zone operative [31].	15
2.3	Dettaglio turni per il periodo dal 26/10/2022 al 26/10/2022.	18
2.4	Dettaglio permanenze nelle aree di controllo dal 26/10/2022 al 26/10/2022.	18
2.5	Allarmi fermate per il periodo dal 26/10/2022 al 26/10/2022.	19
3.1	Struttura della base dati relativa ai Giri LeO	23
3.2	Struttura base dati relativa alle letture nei 12 mesi	24
3.3	Struttura base dati relativa alle pesate sui semirimorchi	25
3.4	Struttura base dati relativa alle pesate all'impianto di destinazione	25
3.5	Struttura base dati relativa alle letture di 5 anni	26
3.6	Struttura base dati relativa ai pesi di 5 anni	26
3.7	Struttura base dati relativa alle letture associate ai relativi pesi nei 5 anni.	29
4.1	Istogrammi relativi alle letture effettuate dai giri di raccolta rispettivamente per rifiuto Vegetale e Secco.	39
4.2	Istogrammi relativi alla volumetria raccolta dai giri di programmati rispettivamente per rifiuto Vegetale e Secco.	40
4.3	Istogrammi relativi ai kg di rifiuto raccolti dai giri di programmati rispettivamente per rifiuto Vegetale e Secco.	40
4.4	Grafici generati dalla funzione <i>fitdist()</i> sui dati relativi alle Letture del rifiuto Vegetale.	42
4.5	Grafici generati dalla funzione <i>fitdist()</i> sui dati relativi alla volumetria del rifiuto Secco.	43
4.6	Grafici generati dalla funzione <i>fitdist()</i> sui dati relativi ai kg raccolti di rifiuto Vegetale.	43
4.7	Grafico che mette in relazione letture e media di densità per il rifiuto Secco.	45
4.8	Grafico che mette in relazione due serie storiche (Letture e densità mensile) per il rifiuto Vegetale.	49
4.9	Grafico che mette in relazione due serie storiche (Letture e densità mensile) per il rifiuto Secco non riciclabile.	49

4.10	Grafico che mette in relazione due serie storiche (Letture e media volumetria ensile) per il rifiuto Vegetale.	50
4.11	Grafico che mette in relazione due serie storiche (Letture e media volumetria mensile) per il rifiuto Secco non riciclabile.	50
4.12	Grafico che mette in relazione due serie storiche (Letture e Kg) per il rifiuto Vegetale.	51
4.13	Grafico che mette in relazione due serie storiche (Letture e Kg) per il rifiuto Secco non riciclabile.	51
4.14	Grafico che mette in relazione due serie storiche (densità ed % esposizione) per il rifiuto Vegetale.	52
4.15	Grafico che mette in relazione due serie storiche (densità e % esposizione) per il rifiuto Secco non riciclabile.	52
4.16	Decomposizione della serie storica relativa alle letture del rifiuto Vegetale.	54
4.17	Decomposizione della serie storica relativa alle letture del rifiuto Secco non riciclabile.	54
4.18	Decomposizione della serie storica relativa ai kg del rifiuto Vegetale.	55
4.19	Decomposizione della serie storica relativa ai kg del rifiuto Secco non riciclabile.	56
4.20	Decomposizione della serie storica relativa alla media di volumetria del rifiuto Vegetale.	57
4.21	Decomposizione della serie storica relativa alla media di volumetria del rifiuto Secco non riciclabile.	58
4.22	Istogramma distribuzione letture rifiuto Secco non riciclabile.	59
4.23	Dati osservati, valori stimati sulla base del modello e retta stimata ai minimi quadrati rifiuto Vegetale	64
4.24	Dati osservati, valori stimati sulla base del modello e retta stimata ai minimi quadrati rifiuto Secco	64
5.1	Base algoritmo operativo utilizzato da Contarina SpA.	66
5.2	Parte nuovo algoritmo operativo utilizzato da Contarina SpA per il rifiuto Secco.	67

Giulio Zanatta: *Analisi e gestione dei dati per il servizio di raccolta rifiuti porta a porta presso Contarina SpA e sviluppo di un modello predittivo*, Tesi di laurea, ©
Febbraio 2023.

Capitolo 1

Introduzione

In questo capitolo vengono introdotte l'azienda ospitante dove il laureando ha svolto l'attività di *stage*, e si fornisce una descrizione dello *stage* stesso.

1.1 L'azienda

Contarina SpA [10] è una società che si occupa della gestione dei rifiuti nei 49 comuni aderenti al Consiglio di Bacino Priula, situati all'interno della provincia di Treviso, attraverso un sistema integrato che considera il rifiuto dalla produzione, alla raccolta, al trattamento e recupero, producendo un impatto positivo sia sulla natura che sulla vita dei cittadini. Gli obiettivi quotidiani dell'azienda consistono nel raggiungimento di una percentuale sempre più elevata di raccolta differenziata, che già si stanZIA all'85%, nella riduzione della quantità dei rifiuti prodotti, particolarmente per quanto riguarda la percentuale non riciclabile, nell'innalzamento della qualità del materiale riciclabile raccolto e del servizio offerto, per ottimizzare il rapporto fra i costi e i benefici nonché promuovere azioni di riciclo e riuso.

Nel rispetto delle norme europee e italiane, la società gestisce i rifiuti attraverso un modello ispirato al paradigma dell'economia circolare, dove ogni bene e servizio è concepito al fine di essere ecosostenibile. L'ottica circolare mira a minimizzare gli sprechi e a reimpiegare quanto apparentemente sembra scarto. Contarina è una società che si posiziona da anni ai vertici europei in termini di raccolta differenziata e rappresenta un esempio di realtà pubblica all'avanguardia nei servizi e nei risultati, operando con il fine comune cioè la tutela dell'ambiente. La strategia aziendale si basa sul continuo miglioramento dei servizi richiesti: la volontà è quella di essere un'impresa innovativa e tecnologicamente all'avanguardia, riducendo l'impatto ambientale che ne deriva. Nel 2020 Contarina ha raccolto e gestito un totale di 229.447 tonnellate di rifiuti urbani e assimilati agli urbani, di cui l'89% è stato raccolto in modo differenziato, e tutti i 49 comuni hanno presentato una soglia minima dell'85% almeno [5]. In Figura 1.1 viene illustrato negli anni come la percentuale di rifiuti raccolti da Contarina Spa in maniera differenziata aumenti. La controparte invece, stabilisce che la produzione di rifiuto non riciclabile pro-capite non è inferiore ai 27 kg/ab.*anno per ogni comune [22].

L'utente di Contarina ha prodotto in media 413,34 kg di rifiuto all'anno. Umido e vetro-plastica-lattine sono le tipologie maggiormente prodotte: per queste categorie le frequenze di raccolta sono più alte, al fine di garantire ai cittadini il miglior servizio

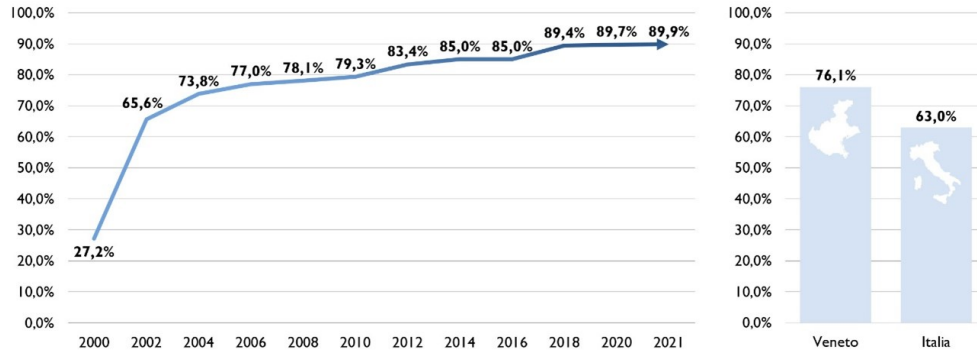


Figura 1.1: Percentuale di raccolta differenziata, dati Contarina, [22], [6]

possibile [5].

Per aiutare a sostenere l'ambiente, inoltre, la società Contarina fornisce il servizio di EcoCentri organizzato in modo da integrare e completare il funzionamento del modello di gestione, consente agli utenti di portare quei rifiuti che non possono essere conferiti, per dimensione o altre caratteristiche, nel normale circuito di raccolta porta a porta e che diversamente andrebbero a compromettere la qualità dei materiali raccolti.

1.2 Proposta di stage

Lo *stage* proposto consiste nella comprensione e nell'analisi del servizio erogato dall'azienda per la raccolta dei rifiuti porta a porta, con la finalizzazione allo sviluppo di una proposta di riorganizzazione dell'attuale piano in atto e di un modello predittivo. L'attuale sistema che Contarina SpA utilizza per organizzare la raccolta dei rifiuti porta a porta si presenta come un sistema statico, in pratica l'azienda per erogare il servizio in una determinata area si presta al calcolo delle risorse necessarie per soddisfare tutte le utenze presenti nell'area. Questo calcolo viene effettuato tenendo in considerazione molti parametri come ad esempio il tempo medio impiegato per effettuare il percorso, quante utenze sono presenti nell'area.

Questo non tiene conto di eventuali stagionalità legate a determinate tipologie di rifiuto, e quindi potrebbe non garantire l'efficienza del servizio ottimale durante tutto l'anno, in quanto quest'ultimo calcola le risorse necessarie in base a dei parametri statici.

Successivamente all'analisi e alla rilevazione di eventuali stagionalità, si richiede lo sviluppo di un modello predittivo di determinati parametri di cui si vuole conoscere al meglio la previsione futura.

1.3 Descrizione dello stage

Lo *stage* proposto consiste nello sviluppo di una proposta di riorganizzazione e di un modello predittivo ai fini di migliorare l'efficienza del servizio di raccolta rifiuti porta a porta dell'azienda.

La proposta di riorganizzazione prevede di rendere meno statico il calcolo delle risorse necessarie per effettuare il servizio, in particolare si vogliono considerare i fattori stagionali, i quali non sono considerati dal calcolo attuale, rischiando di portare il

servizio ad un sovraccarico o mancanza di risorse per il suo normale svolgimento. L'algoritmo attualmente utilizzato, infatti, non tiene conto di fattori come un eventuale andamento stagionale della produzione di determinate tipologie di rifiuto, oppure se effettivamente le utenze espongano sempre il contenitore nel giorno di raccolta programmato per quel determinato rifiuto.

STIVA	ADDEBITI	FATTURE C	CORRONE	DEBITO	PERMANENZA (EUI)	PERMANENZA (MIG. TAVOLA)	ESTEN. TOTALE	% ESTENSIONE	ESTEN. ESTERNA	ESTEN. INTERNA	TEMPO TAVOLA (EUI)	AMMORTAMENTO (EUI)	TEMPO TAVOLA (MIG.)	PERMAN. (EUI)	OPERAZIONI CREDITAZIONE	TEMPO TAVOLA (MIG. TAVOLA)	AMMORTAMENTO (EUI)	PERCENTUALE PERCENTUALE	ESISTE
ROSA	4.471	2	ARCADE	VEGETALE	240	1	204	25	526	7.60	0.05	3	0.40	0.00	0.25	5.50	30	93	8.01
GALLA	7.843	2	ARCADE IN PAVI	VEGETALE	240	7	1.532	74	7.60	0.00	2.70	0.40	0.00	0.00	0.00	6.10	42	91.3	4.02
GALLA	11.173	2	CALABRIMA	VEGETALE	240	1	1.372	25	4.00	7.60	0.00	3	0.75	0.00	0.25	6.02	30	91	4.00
RO	9.332	2	CANDIDA SUL. TALE	VEGETALE	240	1	2.030	35	771	7.60	1.33	3	0.50	0.00	0.25	6.02	30	91	5.5
ROSA	1.344	2	CASER	VEGETALE	240	1	1.033	25	546	7.60	0.01	3	0.25	0.00	0.25	5.10	30	91	4.02
ROSA	5.362	3	GAGNER DEL MONTELO	VEGETALE	240	1	236	25	924	7.60	0.06	3	0.75	0.00	0.25	5.34	30	90	6.03
ROSA	4.353	3	MAJORA DEL. PAVI	VEGETALE	240	1	1.033	25	360	7.60	0.03	3	0.75	0.00	0.25	4.27	30	90	2.8
BLU	4.258	3	MINISTRI DEL TIRUGO	VEGETALE	240	1	1.033	25	360	7.60	0.03	3	0.75	0.00	0.25	4.27	30	90	2.8
VERDE	4.401	3	MICHIANI	VEGETALE	240	1	960	35	1.401	0.00	0.00	0.00	0.00	0.00	0.25	4.46	30	90	1.4
ROSA	8.841	4	NEPOTIA DELLA BATTAGLIA	VEGETALE	240	1	362	25	127	7.60	0.06	3	0.41	0.00	0.25	4.86	25	94	1.0
VERDE	9.196	2	PACIN	VEGETALE	240	1	2.463	25	1.033	0.00	0.00	0.00	0.00	0.00	0.25	4.21	30	90	1.4
CELESTE	52.495	2	PONCARIO VENETO	VEGETALE	240	1	1.175	25	41	7.60	0.00	3	0.31	0.00	0.25	5.42	30	90	2.5
CELESTE	5.255	3	POTEGIA VENETO	VEGETALE	240	1	391	35	337	7.60	0.06	3	0.50	0.00	0.25	5.45	30	94	0.4
ROSA	36.806	2	PREGACIANO	VEGETALE	240	1	2.149	25	932	7.60	0.00	3	0.20	0.00	0.25	5.50	30	93	6.8
VERDE	1.854	2	QUANTO DI REVERSI	VEGETALE	240	1	1.854	25	384	7.60	0.02	3	0.25	0.00	0.25	5.27	30	90	2.2
BLU	16.283	3	FORNACE	VEGETALE	240	1	1.740	35	609	7.60	0.03	3	0.41	0.00	0.25	4.77	30	94	0.3
BLU	10.032	3	SANFONICO DI CALALATA	VEGETALE	240	1	1.024	35	449	7.60	0.03	3	0.41	0.00	0.25	4.71	30	91	3.2
ROSA	16.147	2	SARACINO	VEGETALE	240	1	1.162	25	1.162	0.00	0.00	0.00	0.00	0.00	0.25	4.46	30	91	4.0
ROSA	32.620	3	SEVERIANO	VEGETALE	240	1	1.297	35	454	7.60	0.33	3	0.75	0.00	0.25	4.27	30	93	3.5
ROSA	19.551	3	SIEGACINA	VEGETALE	22	2	1.105	35	397	7.60	0.06	3	0.41	0.00	0.25	4.86	30	90	2.7
ROSA	14.652	1	SIEGACIN - DONA 3	VEGETALE	240	1	7.157	25	7.60	0.05	3	0.41	0.00	0.00	0.25	4.71	30	91	2.1
CELESTE	58.009	2	VALLORIA	VEGETALE	240	1	2.323	25	360	0.00	0.00	3	0.25	0.00	0.25	5.74	30	92	4.0

Figura 1.2: Algoritmo utilizzato per il calcolo delle risorse in uso tutt'oggi da Contarina SpA.

Come si può notare da Figura 1.2, il calcolo è impostato per comune e per tipologia di rifiuto, quindi non tiene conto di una possibile variazione nei mesi che compongono l'anno e di conseguenza di un possibile aumento o calo del numero di risorse necessarie.

1.3.1 Obiettivi

Nella Tabella 1.1 vengono elencati gli obiettivi previsti dall'attività di *stage*. Vengono indicati gli obiettivi *Obbligatorî* con la lettera "O" presente nel codice obiettivo, gli obiettivi *Desiderabili* con la lettera "D" mentre gli obiettivi *Facoltativi* con la lettera "F".

1.3.2 Pianificazione del lavoro

L'attività di *stage* è stata pianificata per esser volta in un totale di 300 ore lavorative, escluse le festività, suddivise in 8 ore giornaliere. La Tabella 1.2 riassume le attività preventivate e le relative ore assegnate.

In Figura 1.3 viene mostrato il diagramma di Gantt relativo alla pianificazione della attività su base settimanale.

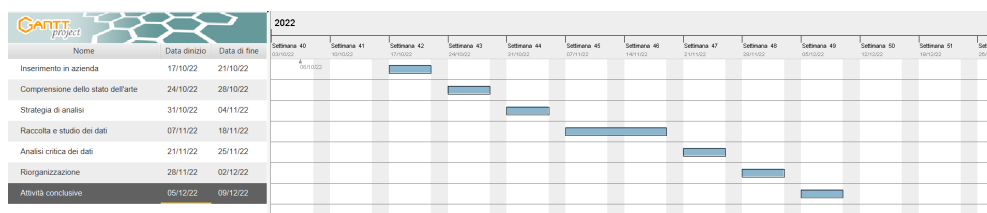


Figura 1.3: Diagramma di Gantt relativo alla pianificazione dell'attività di stage.

Tabella 1.1: Tabella degli obiettivi.

Descrizione obiettivo	Codice obiettivo
Comprensione dell'architettura As Is	O01
Esecuzione analisi dati e caratterizzazione delle variabili	O02
Elaborazione di una proposta di riorganizzazione operativa del servizio PAP	O03
Esecuzione Value Stream Map	D01
Individuazione del set di dati da passare al data wharehouse	D02
Valutazione dei sistemi informatici attualmente in uso e spunti di miglioramento	F01

Tabella 1.2: Tabella delle attività pianificate.

Descrizione attività	Ore preventivate
Inserimento in azienda	38
Comprensione dello stato dell'arte	38
Strategia di analisi	38
Raccolta e studio dei dati	76
Analisi critica dei dati	38
Riorganizzazione	38
Attività conclusive	38

Capitolo 2

Comprensione dello stato attuale

2.1 Produzione di rifiuti

Lo sviluppo sociale ed economico ha causato, a livello internazionale e a livello nazionale, un aumento della produzione di rifiuti urbani. In territorio Europeo, nell'anno 2020, ogni cittadino ha prodotto mediamente 505kg di rifiuti, aumentando dell'1% rispetto al 2019 e 27.7% rispetto al 2005, producendo un totale di 225.7 milioni di tonnellate di rifiuti urbani.

In territorio italiano invece la produzione di rifiuti si attesta intorno ai 28,9 milioni di tonnellate nell'anno 2020, dove ogni cittadino italiano produce circa 488kg di rifiuti all'anno. Nello specifico nel centro Italia un abitante produce mediamente 548kg, il valore più alto a livello nazionale; al nord la produzione media è di circa 518kg per abitante, mentre al sud è di 445kg per abitante.

L'incremento della produzione si registra nel nord Italia, con una stima di 14,4 milioni di tonnellate di rifiuti, mentre è in calo al centro con circa 6,6 milioni di tonnellate e al sud con 9,1 milioni di tonnellate [7].

Infatti, tra il 2018 e il 2019, 13 regioni italiane, in particolare quelle meridionali, hanno fatto rilevare un calo della produzione dei rifiuti urbani al contrario di quanto avvenuto nelle regioni settentrionali. Rispetto agli anni precedenti si presenta comunque un lieve calo dello 0.3% degli stessi rispetto al 2018. Dato l'aumento della produzione di rifiuti, viene necessario un aumento della raccolta differenziata di quest'ultimi che vede raddoppiare rispetto al 2008 la percentuale di copertura del servizio riuscendo a coprire il 61.3% della produzione nazionale.

A livello regionale vi è un aumento significativo della raccolta differenziata dei rifiuti urbani, riuscendo a superare nel 2019 l'obiettivo di copertura del 65% per ben 8 regioni [8].

La regione Veneto riesce a superare ormai da molti anni l'obiettivo del 65% previsto della normativa nazionale, collocandosi al primo posto tra le Regioni italiane. Tutte le province oltrepassano l'obiettivo di raccolta differenziata del 65% previsto dal d.lgs. 152/06 per il 2012 e Treviso con 88,3%, Belluno con il 76,8% e Vicenza con 78,2% superano anche il 76% previsto per il 2020 dagli scenari del Piano Regionale Rifiuti [9]. I numeri per la quantità di rifiuti urbani raccolti in modo differenziato dimostrano come le province del Veneto si siano impegnate per differenziare quantitativi sempre più elevati di rifiuti, raggiungendo ottime percentuali di raccolta differenziata che nel 2020 permettono al Veneto di arrivare al 76,1%, il valore più alto a livello nazionale [7]. Il sistema di raccolta rappresenta uno degli aspetti fondamentali nella gestione dei

rifiuti urbani, da un lato perché influenza la quantità e la qualità dei rifiuti che vengono intercettati ed avviati a recupero o smaltimento dall'altro perché incide sul costo del servizio di raccolta.

2.2 Classificazione del territorio servito

Contarina SpA effettua la raccolta porta a porta in 49 comuni nella Provincia di Treviso, territorio con 555.000 abitanti su una superficie totale di 1.300 km^2 con differenti tipologie di aree [20].

La raccolta dei rifiuti urbani viene svolta su tutto il territorio mediante il passaggio su aree pubbliche o ad uso pubblico. Le aree servite sono classificate in base alla loro complessità urbana e identificate come:

- * Zone a servizio standard, ovvero sono caratterizzate da uno sviluppo urbano orizzontale con condomini e utenze non complesse, aventi spazi di proprietà interni ed esterni;
- * Zone urbanisticamente complesse, ovvero sono caratterizzate da uno sviluppo urbano prevalentemente verticale, con alta densità abitativa e presenza di condomini e utenze complesse, con spazi di proprietà ridotti o assenti per il posizionamento dei contenitori;
- * Zone a bassa densità abitativa, ovvero sono zone caratterizzate da uno sviluppo urbano esclusivamente orizzontale, con edifici sparsi, coincidente generalmente con zone montane dove le utenze sono spesso utilizzate stagionalmente;
- * Centri storici, il territorio include i centri storici di Treviso, Asolo, Castelfranco e Montebelluna.



Figura 2.1: Mappa del territorio servito [20]

Il territorio servito da Contarina presenta caratteristiche urbanistiche e morfologicamente differenti. La suddivisione in zone, come mostrato in Figura 2.1, e la

loro classificazione ha permesso, quindi, di individuare le soluzioni più adeguate a organizzare il servizio in modo più efficace e sostenibile.

Le frazioni di rifiuto sono omogenee, mentre le tipologie dei contenitori e le frequenze di raccolta sono legate alle caratteristiche del territorio.

2.3 Volumetrie

Alle utenze vengono forniti contenitori per ogni tipologia di rifiuto e con una dimensione adeguata in base alla classificazione della zona.

Si distinguono contenitori per volumetria (come riportato in [20], [6]):

- * 22 litri per il contenitore dell'umido;
- * 30 litri per il contenitore di secco, carta e VPL;
- * 120 litri / 240 litri per il contenitore di umido, vegetale, secco, carta e VPL;
- * 360 litri per il contenitore di secco, carta e VPL.

Le volumetrie sono attribuite in base alla classificazione in cui si trova l'utenza:

- * Utenze standard con una capacità di 22 litri per umido e 120 litri per VPL, secco, carta, vegetale;
- * Utenze con spazi ridotti con capacità di 22 litri per umido e 30 litri per VPL, secco, carta, vegetale;
- * Utenze centri storici, pari alla situazione di utenze con spazi ridotti.

2.4 Il parco mezzi

Per svolgere il servizio di raccolta, Contarina SpA è dotata di un parco mezzi ampio e differenziato. Questo garantisce un'efficienza elevata per quanto riguarda l'ottimizzazione dei tempi di lavoro e la sicurezza.

Il parco mezzi si compone di 572 mezzi, di cui 280 sono mezzi a vasca utilizzati nella raccolta porta a porta, 30 semirimorchi utilizzati per il trasporto dei rifiuti raccolti durante il servizio. Il numero di mezzi a vasca in servizio da lunedì a venerdì varia da 230 a 250.

2.4.1 Vaschette

Per quanto riguarda il servizio porta a porta, vengono impiegati mezzi di piccole e grandi dimensioni, che svolgono attività diverse e complementari: i più grandi si occupano di trasporto, mentre i più piccoli svolgono le raccolte. Quest'ultimi sono di tipologie diverse, infatti vi sono i mezzi a singola vasca con allestimento per il carico dei rifiuti che può essere posizionato a lato dell'automezzo oppure posteriore.

Questi mezzi presentano una struttura differente tra loro dovuta al tipo di allestimento, che conferisce una diversa portata utile in base a quest'ultimo. La portata utile viene calcolata basandosi sulla massa massima a carico ammissibile del veicolo in

servizio meno la tara dello stesso, quindi pesato a carico vuoto.

La portata utile subisce una maggiorazione del 5% per quanto riguarda i veicoli a diesel [28].

Solo una tipologia di automezzo è allestita con doppia vasca, con la possibilità di carico contemporaneo in entrambe rispettivamente uno laterale e uno posteriore, ed è impiegata nelle località più difficili da gestire come il centro storico della città di Treviso. Questo automezzo permette, avendo due vasche, di raccogliere una o più tipologie di rifiuto.

Durante il giro di raccolta, ogni svuotamento viene registrato tramite la lettura di un trasponder UHF o LF presente sul contenitore, nel caso di rifiuto secco non riciclabile e rifiuto vegetale in tutti i 49 comuni del bacino mentre per tutte le tipologie di rifiuto nel centro urbano del comune di Treviso e nei 4 centri storici. La velocità delle operazioni di svuotamento garantisce tempi di sosta molto brevi [28], limitando eventuali ostacoli al normale flusso del traffico stradale.

2.4.2 Semirimorchi

I semirimorchi sono i mezzi impiegati per completare il trasporto del rifiuto raccolto, possono riguardare un intero comune o più comuni. Questi mezzi hanno una capacità di $52 m^3$ e sono posizionati nel territorio in specifiche aree scelte in base a vincoli di viabilità, decoro ed accessibilità (previa autorizzazione comunale) in punti possibilmente baricentrici rispetto alle zone di raccolta, così ogni volta che un mezzo piccolo ha completato il carico, va a conferire il suo contenuto all'interno del semirimorchio.

Durante il servizio di raccolta, una volta raggiunta la capacità degli automezzi a vasca, l'operatore interrompe il servizio e chiede al sistema di navigazione di essere indirizzato verso il semirimorchio per lo scarico del rifiuto.

Dopo aver eseguito lo scarico, l'operatore chiede al sistema di essere guidato al punto in cui aveva interrotto il servizio e, da lì, riprende il percorso pianificato. L'attività di svuotamento dei mezzi a vasca nel semirimorchio avviene in media dalle 2 alle 4 volte a turno in base alla tipologia del mezzo, alla tipologia di rifiuto raccolto e in base alla porzione di territorio servito.

Il sistema di raccolta con scarico intermedio vige per tutte le frazioni di rifiuto tranne l'umido, che viene conferito alla fine del turno presso l'impianto di Lovadina di Spresiano o presso quello di Trevignano.

2.5 Strumenti utilizzati

Dal 01/01/2017, entra in vigore il Piano Operativo 38 (PO38) [31] e con esso la modifica contrattuale dell'orario di lavoro, il totale ore settimanale è stato modificato da 36 ore a 38 ore.

Questo ha portato alla modifica degli orari dei turni di lavoro, per quanto riguarda il servizio di raccolta rifiuti porta a porta, da 7,15 ore a 7,36 ore.

Gli obbiettivi imposti dal PO38 vedono al miglioramento del servizio con le nuove disposizioni contrattuali; infatti, si punta ad una ottimizzazione dei giri di raccolta dimensionandoli per ottenere un rientro del mezzo in orario ordinario e un numero di risorse necessarie ridotte. Tali disposizioni permettono di impiegare più tempo utile netto nel servizio di raccolta porta a porta.

2.5.1 LeO (Localizzazione e Organizzazione)

Il progetto LeO [32] nasce per la progettazione del servizio di raccolta porta a porta della flotta aziendale con l'obiettivo di conoscere in tempo reale la posizione dei mezzi (storico giornaliero compreso), pianificare l'attività di ogni mezzo attraverso la navigazione e permettere agli operatori di completare correttamente qualsiasi percorso di raccolta.

È un sistema integrato che si basa su due strumenti differenti:

- * LeOWeb, centrale di controllo che si occupa di pianificazione e controllo.
- * LeOCar, navigatore installato a bordo mezzo che fornisce indicazioni per eseguire l'attività.

Il progetto PO38 ha permesso di pianificare la prima fase di progettazione dei giri di raccolta, prevedendo il calcolo del numero di risorse necessarie e standardizzando i percorsi raccolta.

Una volta conclusa la corretta divisione del territorio è necessario il rilievo del percorso di raccolta con l'aiuto dell'autista che ne è titolare. La stesura tiene conto delle strade principali, dei punti di maggiore traffico e di tutte le informazioni necessarie (le quali il traffico, strade particolari, retromarce ecc.) per eseguire correttamente il servizio.

Nella fase di digitalizzazione del percorso, esso viene arricchito con delle specifiche per una corretta navigazione ed un corretto servizio. Infatti, vengono indicati i tratti dove la raccolta va eseguita sul lato destro o sinistro piuttosto che su entrambi i lati, sensi di marcia, possibilità o non di effettuare determinate manovre con il mezzo, in modo da permettere all'operatore di completare il percorso assegnato senza generare disservizi e in totale sicurezza anche se lo esegue per la prima volta.

Per poter gestire la navigazione all'interno di aree private il sistema utilizza uno stradario di proprietà [32].

Il sistema di navigazione, LeOCar ha la funzione di co-pilota a bordo mezzo per il supporto nell'esecuzione dei percorsi di raccolta porta a porta. Si tratta di un navigatore che fornisce sia le indicazioni stradali per la conduzione del mezzo sia le informazioni operative necessarie per eseguire correttamente l'attività di raccolta. La navigazione dei mezzi avviene per mezzo di un'interfaccia hardware composta da una blackbox che permette la localizzazione GPS del mezzo e da un PC che gestisce l'interfaccia di navigazione LeOCar.

La centrale operativa permette la consultazione di dati in tempo reale e di dati storici. Sono quindi previste due sezioni separate per gestire queste informazioni, una sezione per le informazioni in tempo reale e una sezione per le informazioni relative allo storico.

La sezione delle informazioni in tempo reale riguarda dati come la verifica della posizione dei mezzi in tempo reale, l'interazione con i mezzi sul territorio (messaggi, obiettivi...), informazioni in tempo reale sullo stato delle raccolte e una reportistica in tempo reale.

La sezione delle informazioni riguardanti lo storico comprendono, invece, la verifica delle tracce GPS registrate, conferimenti eseguiti ai semirimorchi, rendicontazione dell'attività eseguita e una reportistica su archi temporali specifici [29].

2.5.2 ECOS

ECOS è una suite software che permette la gestione interna della rendicontazione dei dati. Contarina lo adotta, in ambito di raccolta porta a porta, per generare documenti necessari per lo svolgimento dell'attività di raccolta e per raccogliere informazioni sulle pesate dei mezzi, sia semirimorchi che vaschette.

Il software permette di generare Ordini di Servizio (ODS), cioè dei moduli che specificano l'attività che si va a svolgere, in questo caso il servizio porta a porta.

In particolare, un ODS contiene i dati relativi all'attività che si attua, cioè il tipo di servizio e per quale tipologia di rifiuto si effettua, l'automezzo utilizzato, l'operatore e la durata effettiva del turno. Specifica inoltre chi è il produttore del rifiuto che si va a raccogliere, inteso come zona cioè il comune di interesse, il trasportare e il destinatario del rifiuto raccolto, cioè Contarina nel nostro caso.

SGQ (Segnalazione Gestione Qualità), modulo generato da ECOS che va compilato dall'operatore stesso nel caso si incontrino difformità durante il giro di raccolta. Quest'ultime possono comprendere la presenza di un rifiuto non conforme nel contenitore, un contenitore troppo pieno, contenuto del contenitore sparso nelle vicinanze, obbligo di utilizzare un sacchetto biodegradabile per contenere il rifiuto umido all'interno del proprio contenitore riservato e doppi svuotamenti. Dopo aver compilato il modulo SGQ l'operatore applica un adesivo rosso (presente nel modulo) che andrà ad informare l'utenza di segnalazione.

In quest'ultimo caso, vi è da porre attenzione ad un doppio svuotamento del contenitore del rifiuto secco non riciclabile. Lo svuotamento di questa tipologia di rifiuto è a pagamento ed ogni utenza è disposta (per contratto) ad un numero limitato di svuotamenti, quindi in caso di doppio svuotamento verrà messa in conto all'utenza la vuotata in eccesso.

Se il doppio svuotamento viene invece effettuato per il contenitore di VPL, carta e cartone, umido, viene generato l'SGQ e viene segnalata l'utenza.

2.5.3 DBW

L'azienda predispone di un software per la gestione delle utenze, chiamato DBW, dove sono contenute le anagrafiche dei soggetti fisici e le informazioni relative alle utenze.

Inoltre, contiene informazioni di geolocalizzazione dei civici da soddisfare per quanto riguarda il servizio di raccolta porta a porta.

2.6 Pianificazione ed esecuzione del servizio PaP

2.6.1 Pianificazione

Per quanto riguarda la pianificazione, essa si basa su due macro-strumenti, una carta d'identità per ogni giro di raccolta e l'algoritmo per il calcolo delle risorse necessarie, ma tiene conto anche dell'esperienza lavorativa degli operatori di servizio.

Per ottenere una carta d'identità per ogni giro di raccolta, vi è una procedura informatica per l'estrazione dei dati da ogni singolo giro, utilizzando le cartografie in AutoCAD e le banche dati relative alla gestione delle utenze.

I dati di importanza che vengono rilevati (per ogni singolo giro) sono:

- * Numero totale di utenze;

- * Numero contenitori per ogni tipologia di raccolta;
- * Tipologia dei contenitori.

La creazione della carta d'identità di un determinato giro di raccolta viene impostata procedendo per step [31]:

1) Creazione di un poligono dell'area di interesse che definisce la zona in cui si sviluppa ogni singolo giro di raccolta, mettendo in evidenza con colori diversi ogni giro, delimitando la zona del singolo giro con una linea nera ed evidenziando il civico dell'utenza con un pallino giallo;

2) Caricamento dalla banca dati del gestionale, le utenze che sono associate ad ogni civico e di conseguenza i contenitori associati al contratto di quell'utenza con relative volumetrie;

3) Estrazione dei dati per ogni singolo giro di raccolta e per ogni tipologia di materiale raccolto, ottenendo un numero complessivo di giri per area per ogni tipologia di rifiuto.

Notare che i dati prelevati nel punto 2) sono statici, non dinamici, di conseguenza se subiscono una variazione, e quindi si necessita di un aggiornamento, bisognerà effettuare un'ulteriore estrazione e relativo collegamento. Ogni 12 mesi è comunque previsto un aggiornamento per eventuali variazioni.

Per ogni giro viene considerato inoltre il confronto tra la lunghezza del percorso in metri e il totale dei contenitori da vuotare in quel determinato giro, in modo da poter ottenere una densità per quel giro.

I dati riguardanti l'esposizione dei contenitori sono stati rilevati analizzando un giro di raccolta in un determinato comune nel periodo da novembre 2016 fino ad Aprile 2017, evidenziando che i contenitori di carta, vetro-plastica-lattine (VPL), secco, umido e vegetale hanno un'esposizione media di rispettivamente 50%, 45%, 14%, 65%, 35% [28].

L'algoritmo per il calcolo delle risorse necessarie per il servizio di raccolta porta a porta presenta determinate specifiche nel modello. Infatti, vengono prese in analisi tutte le procedure operative del servizio, vengono studiate nel dettaglio le varie situazioni operative e vengono individuati tutti gli elementi possibili rilevabili e traducibili in linguaggio informatico.

In base a queste specifiche vengono individuate delle variabili costanti dell'algoritmo, cioè fisse e non condizionabili, che sono il tempo necessario al trasferimento dalla sede di Lovadina al comune da servire e viceversa, le tempistiche di scarico verso il semirimorchio e la pausa usufruibile dagli operatori in servizio fissata a 30 minuti.

Vengono individuati, inoltre, ulteriori parametri variabili e non costanti, come la percentuale di esposizione dei contenitori soggetti alla raccolta differenziata per tipologia di rifiuto e densità delle utenze in un determinato comune, cioè la lunghezza delle strade da servire in un comune in rapporto al numero di utenze da servire ottenendo una media della distanza tra un contenitore ed un altro in quella zona. Quest'ultima variabile incide sulla produzione operativa, in quanto per quest'ultima si intende il numero totale di contenitori vuotati in un'ora netta di servizio e di conseguenza dipende dalla densità delle utenze presenti dove si effettua il servizio.

Nell'analisi della produzione operativa, viene calcolato il ritmo di lavoro dei contenitori vuotati in un dato intervallo di tempo cioè il numero dei contenitori vuotati in un'ora al netto delle tempistiche di trasferimento verso il comune da servire e dal comune

dove è stato appena effettuato il servizio, scarico e pausa.

Si denota come la produzione operativa sia influenzata da fattori come il numero di utenze, la distanza tra i contenitori da raccogliere, la conformazione del territorio, la densità urbana e zone con alta densità di utenze.

La fase finale della pianificazione comprende un'analisi per distribuire le raccolte nei giorni della settimana. Come detto in precedenza, ad ogni comune sono associati dei giri di raccolta per ogni tipologia di rifiuto, e per ogni giro, il numero di risorse necessarie per eseguirlo [28].

Per quanto riguarda le frequenze di raccolta esse variano da zona a zona [33], infatti per le zone a servizio standard e a bassa densità abitativa, le frequenze di raccolta per secco non riciclabile e carta sono ogni 15 giorni, mentre per l'umido 2 volte a settimana, ed infine 1 volta a settimana per VPL e vegetale.

Nelle zone urbanisticamente complesse per il contenitore del secco non riciclabile e carta la frequenza varia da 2 volte a settimana ad 1 volta a settimana. Per il contenitore dell'umido la frequenza è 3 volte a settimana, mentre per VPL varia da 3 volte a settimana a 2 volte a settimana. Infine, per il contenitore del vegetale la frequenza è 1 volta a settimana.

La distribuzione settimanale viene pianificata tenendo come obiettivo la linearità della forza lavoro disponibile. Un eccesso di quest'ultima causerebbe un aumento di costi inutile ed evitabile; quindi, si cerca di mantenere una continuità lineare, evitando improvvisi cambiamenti come cali o aumenti significativi. Si tiene conto anche dell'assenteismo, dovuto ad assenza corte o lunghe (3 settimane), in modo da poter rimpiazzare la forza lavoro che viene a mancare.

La pianificazione del servizio, in particolare, si suddivide in due parti cioè la prima settimana e la seconda settimana rendendo la pianificazione quindicinale. Questa divisione è dovuta al modello Contarina adottato, che dice che è sufficiente effettuare 1 giro di raccolta ogni 2 settimane per i rifiuti quali secco non riciclabile e carta. Se così non fosse, la programmazione del servizio si potrebbe pianificare in una settimana, ma le esigenze risultate dagli studi effettuati non lo permettono.

2.6.2 Esecuzione

Il servizio è organizzato secondo un Piano Operativo che suddivide il territorio in 10 zone, all'interno delle quali il personale lavora in una o più squadre coordinate da un Capo Squadra.

Il Piano Operativo è articolato in circa 1450 giri di raccolta: tutti i giri sono fissi, e seguono un preciso percorso indicato all'operatore da un sistema di navigazione satellitare costruito basandosi sull'esperienza degli operatori con lo scopo di erogare il servizio nelle arterie principali e in luoghi sensibili a fasce orarie in cui potrebbero esserci delle problematiche relative al traffico e/o alla viabilità.

L'orario di servizio è articolato in turni da 7 h 36 min su 5 giorni la settimana (da lunedì a venerdì) come da PO38, suddividendo gli orari per zone [28]:

- * 05:00 - 12:36 per 6 zone: Rosa, Celeste, Gialla, Blu, Verde e Rossa;
- * 04:30 - 12:06 per 3 zone: Marrone, Viola e Arancione.

In Figura 2.2 mostra il territorio suddiviso per zone e differenziate per colore. Si può notare la zona di colore bianco, che si esula dal programma operativo base in quanto lavora su turni da 9 h 30 min (04:00-13:30) per 4 giorni a settimana e con riposo nella giornata del mercoledì: si tratta di una sperimentazione iniziata nel 2021

nell'area territorialmente più distante dalla sede di Spresiano per limitare l'impatto dei costi di trasferimento da e per la sede stessa. Essa è l'unica zona in cui alcuni giri di raccolta non rispettano il vincolo di rimanere all'interno del singolo Comune.

Il sabato sono pianificati circa 30 turni di raccolta, ma lavorativa solo per gli operatori che prestano servizio nei centri storici di Treviso, Asolo, Castelfranco e Montebelluna. Tutti i turni prevedono una pausa di 30 minuti che l'operatore può gestire liberamente.



Figura 2.2: Suddivisione del territorio in zone operative [31].

2.6.3 Partecipazione ai turni di raccolta

La partecipazione alla raccolta dei rifiuti PAP è importante per poter comprendere al meglio quali parametri influiscono maggiormente sulla corretta esecuzione di un turno di raccolta programmato.

Il primo turno di raccolta al quale lo stagista ha partecipato è stato un giro programmato per il rifiuto Umido:

- * Mezzo: RU317
- * Tipologia rifiuto: Umido
- * Data: 25/10/22

Il mezzo utilizzato è un automezzo a vasca con carico del rifiuto posizionato lateralmente. Il contenitore, non appena l'operatore arriva in prossimità dell'utenza da servire e aziona il freno a mano del mezzo, viene vuotato nell'apposita vasca appendice laterale. Nel caso in cui quest'ultima risulti piena si procede con l'azione di vuotamento nella vasca principale, altrimenti, si continua il giro di raccolta.

Le tempistiche rilevate evidenziano come un giro di raccolta venga influenzato da fattori come problemi interni, traffico, viabilità.

Il giro stesso di raccolta ha subito un ritardo in principio, causa un problema ad un sensore, di circa 30 min. In questo caso, essendo il giro di raccolta per questa

tipologia di rifiuto pianificato con più di 1 risorsa necessaria, gli operatori hanno dovuto organizzarsi per aiutarsi a vicenda.

L'organizzazione tra operatori è molto importante perché permette di svolgere un giro di raccolta in modo esaustivo e completo, anche con il verificarsi di imprevisti. In particolare, durante questo giro, è stata coperta la strada principale da un altro operatore il quale non aveva in programma di percorrerla. L'operatore designato per percorrerla, per il ritardo subito, se l'avesse percorsa in ritardo, quindi fuori pianificazione oraria, si sarebbe imbattuto in traffico e avrebbe causato, molto probabilmente, un rallentamento della viabilità o comunque disagi per gli spostamenti dei cittadini della zona. Questo il motivo principale per cui viene considerato importante il dialogo tra operatori di servizio.

L'operatore, proseguendo con il suo giro di raccolta, può, in base sempre alla sua esperienza maturata sul campo, decidere se intraprendere un rientro anticipandolo nel giro. Questo porta un conflitto con il sistema di navigazione LeOCar, che indica la strada che il mezzo deve percorrere, risolvibile dall'operatore con l'operazione "Salta rientro" un comando da utilizzare quando visivamente non ci sono bidoni da vuotare oppure sono già stati vuotati in precedenza, e permette di proseguire senza accedere nel rientro. Ultimato il giro di raccolta, l'operatore riporta il mezzo in sede a Lovadina. La prima operazione che viene richiesta è la pesa del mezzo carico (oggi era 6800kg), successivamente viene scaricato il mezzo in apposito magazzino. L'operatore, successivamente effettua il lavaggio del mezzo, solitamente viene effettuato nel caso di raccolta di umido o vegetale perché si incorre nel pericolo di corrosione delle vasche causato dalle materie organiche. Viene lavato con un getto d'acqua a pressione soprattutto la vasca e il mezzo in generale.

Dopo aver effettuato il lavaggio, viene fatta una seconda pesatura del mezzo senza carico (oggi era 5150kg) per visionare quanti kg di rifiuto sono stati raccolti. È importante capire che il rifiuto umido non viene scaricato nei semirimorchi, ma direttamente in sede, quando il mezzo ha soddisfatto il giro di raccolta e le relative utenze pianificate. Come ultimo passo, l'operatore si accinge al rifornimento carburante del mezzo e lo parcheggia nel parco mezzi, in retromarcia e sul posto associato al suo mezzo.

Lo stagista ha partecipato ad un ulteriore turno di raccolta, riguardante il rifiuto VPL:

- * Mezzo: RU305
- * Tipologia rifiuto: VPL
- * Data: 26/10/22

Il mezzo utilizzato è un automezzo a vasca con carico del rifiuto posizionato lateralmente, allestimento Novarini, ibrido cioè metano ed elettrico. La vasca appendice è mossa da una batteria e questo velocizza le operazioni di vuotamento del contenitore relativamente di qualche secondo rispetto ad una vasca ad appendice mossa da un impianto idraulico a metano (RU317).

All'operatore è stato assegnato un ODS relativo al giro di Povegliano, ed una volta completato quel giro è stato effettuato un mezzo giro presso Ponzano Veneto con relativo ODS. Per quanto riguarda l'ordine di come vengono affrontati i giri di raccolta, prima è stato completato il giro intero e successivamente il mezzo giro.

Il giro di raccolta viene programmato per soddisfare prima le vie principali e quindi legate a possibili problematiche relative ad orario e traffico, e successivamente i quartieri

e vie secondarie. L'operatore ha percorso e soddisfatto le utenze precedentemente (alle 8.00 circa) di un rientro in via Masetto rispetto a quanto pianificato su LeO, quindi successivamente (alle 10.00 circa) ha deciso di ri-battere quel percorso, per soddisfare la percentuale di completamento di LeO di quest'ultimo. L'operatore doveva soddisfare anche un'utenza, dotata di richiesta apposita per permetter al mezzo in servizio di calcare un suolo privato, dove se vi era il cancello aperto avrebbe dovuto effettuare il carico ma questo caso non si è verificato e l'operatore ha deciso di continuare il percorso.

Un rientro, in particolare visto dal sistema LeOCar, è tecnicamente un tratto di strada lungo circa 500m e senza nodi. Questo implica che il sistema in questione, in presenza di un'utenza avente il cancello in strada ma dotata di richiesta (ad esempio se è anziano), non la valuta come rientro. Di conseguenza se il mezzo passa per l'utenza, non ci possono essere segnalazioni. Il rientro in conclusione, è di fatto un concetto stabile, ma che LeOCar non sempre interpreta al meglio.

Il giro di raccolta si è imbattuto nella generazione di un SGQ, modulo necessario che va compilato dall'operatore stesso nel caso di difformità. Quest'ultime possono comprendere la presenza di un rifiuto non conforme nel contenitore, un contenitore troppo pieno, contenuto del contenitore sparso nelle vicinanze, obbligo di utilizzare un sacchetto biodegradabile per contenere il rifiuto umido all'interno del proprio contenitore riservato e doppi svuotamenti. Dopo aver compilato il modulo SGQ l'operatore applica un adesivo rosso (presente nel modulo) che andrà ad informare l'utenza di segnalazione.

In quest'ultimo caso, vi è da porre attenzione ad un doppio svuotamento del contenitore del rifiuto secco non riciclabile. Lo svuotamento di questa tipologia di rifiuto è a pagamento ed ogni utenza è disposta (per contratto) ad un numero limitato di svuotamenti, quindi in caso di doppio svuotamento verrà messa in conto all'utenza la vuotata in eccesso.

Se il doppio svuotamento viene invece effettuato per il contenitore di VPL, carta e cartone, umido, viene generato l'SGQ e viene segnalata l'utenza.

L'operatore appena ripreso il giro di raccolta ha deciso di effettuare uno scarico nel semirimorchio di Povegliano, affrontando la ripresa a vasca vuota. Successivamente prima di affrontare il mezzo giro a Ponzano Veneto, ha effettuato un secondo scarico. Dovendo iniziare il mezzo giro a Ponzano Veneto, l'operatore ha selezionato su LeOCar il percorso pre-impostato relativo al giro da affrontare.

Quindi, ogni comune ha determinati giri per ogni tipologia di rifiuto da vuotare pre-impostati, compresi i mezzi giri.

2.7 Rendicontazione e raccolta dati del servizio PAP

La rendicontazione e raccolta dati avviene in modo semi-automatico.

Ogni tracciato GPS, in cui è stata utilizzata la navigazione LeOCar, permette di ottenere una rendicontazione dell'attività svolta sul territorio attraverso l'analisi dei tracciati GPS effettuando la verifica tra l'eseguito rispetto al pianificato.

È possibile ottenere report legati all'analisi specifica del singolo mezzo piuttosto che su dati riguardanti l'intera flotta o porzioni di essa, grazie alle tracce GPS rilasciate dai mezzi, che consentono di analizzare le singole attività svolte da essi.

La Figura 2.3 mostra un potenziale report su dati, avente informazioni come inizio e fine del turno, durata dello stesso, lunghezza (*km*), velocità media, velocità massima, percentuale attività turno (cioè in percentuale quanto l'operatore ha utilizzato il tempo

a disposizione nel turno per completare la raccolta), durata effettiva del turno, numero di fermate (senza spegnere il motore) e numero soste con rispettive durate complessive.

Dati generali del turno														Attività				Fermate		Soste			
Mezzo	P. in cart. miss.	Zona	Autista	Comune	Materiale	Missione	Perc. comp. miss.	Perc. comp. serv.	Inizio	Ora Inizio	Fine	Ora Fine	Durata T	Lunghezza	Vel.media	Vel.max	Percentuale attività/turno	Durata attività	N° ferm.	Durata ferm.	N° soste	Durata soste	
26/10/22 Automezzo vasca																							
RU121	SI	VIOLA	ETTA BESSONG NICHOLAS	CASTELFRANCO VENETO	VEGETALE	Castelfranco_VEGETALE 7(955)	100 %	100 %	26/10/2022	04:40:35	26/10/2022	11:11:19	06:30	118.11 km	47 km/h	88 km/h		62 %	04:03	43	02:01	2	00:26
RU122	SI	ARANCIONE	SHKULLAKU MARKET	VEDELAGO	CARTA CARTONE	Ve delago_CARTA 2(733)	87 %	79 %	26/10/2022	04:42:16	26/10/2022	11:43:43	07:01	100.19 km	48 km/h	89 km/h		68 %	04:44	68	02:16	0	00:00
RU123	No								26/10/2022	04:42:34	26/10/2022	12:49:36	08:07	122.26 km	50 km/h	89 km/h		58 %	04:40	89	03:09	1	00:17
RU125	No								26/10/2022	04:39:00	26/10/2022	11:53:35	07:14	151.00 km	46 km/h	89 km/h		71 %	05:06	75	01:53	1	00:14
RU127	SI	VIOLA	BOSCARATO LINO	LORIA	VPL	Loria_VPL 2(984)	99 %	99 %	26/10/2022	04:37:46	26/10/2022	10:57:05	06:19	118.47 km	57 km/h	88 km/h		58 %	03:39	83	02:21	1	00:18
RU128	SI	VIOLA	BARBAN MASSIMILIANO	RESANA	VPL	Resana_VPL 2(1008)	95 %	84 %	26/10/2022	04:37:49	26/10/2022	11:39:10	07:01	143.09 km	51 km/h	88 km/h		69 %	04:50	61	02:10	0	00:00
RU129	SI	VIOLA	DE BENEDETTO FRANCESCO	CASTELFRANCO VENETO	VPL	Castelfranco_VPL 14 (1176)	52 %	58 %	26/10/2022	04:50:13	26/10/2022	12:22:58	07:32	106.13 km	49 km/h	87 km/h		46 %	03:26	97	03:51	1	00:14
RU130	SI	VERDE	STASI ROBERTO	MORGANO	CARTA CARTONE	Morgano_CARTA 2(345)	59 %	91 %	26/10/2022	04:32:18	26/10/2022	12:03:55	07:31	98.09 km	45 km/h	88 km/h		50 %	03:46	118	03:37	1	00:08
RU131	SI	VERDE	LAPONI LUCA	ZERO BRANCO	CARTA CARTONE	Zero Branco_CARTA 1 (646)	67 %	99 %	26/10/2022	05:04:13	26/10/2022	12:57:06	07:52	110.01 km	41 km/h	83 km/h		52 %	04:04	126	03:39	1	00:09

Figura 2.3: Dettaglio turni per il periodo dal 26/10/2022 al 26/10/2022.

Inoltre, possiamo ottenere un report sull'attività giornaliera che riguarderà diversi automezzi vasca, anche in zone differenti, mantenendo la visualizzazione dei dati elencati in precedenza.

Un'altra tipologia di report generabile da LeO è quello riguardante i dati sulla permanenza dei mezzi in determinare aree di controllo come mostrato in Figura 2.4. Quest'ultime sono intese come aree degli impianti oppure come aree semirimorchi. Grazie al segnale GPS di ogni mezzo è possibile capire per quanto tempo quest'ultimo rimane all'interno dell'area di permanenza precedentemente designata.

Codice area	Mezzo	Data	Zona operativa	Comune	Autista	Materiale	Id giro	Missione	Ora Ingresso	Ora Uscita	Durata
26/10/2022 Automezzo vasca											
AIM03	RU184	26/10/2022	ARANCIONE	ISTRANA	MITABLINDA TUDOR	CARTA CARTONE	677	Istrana_CARTA 6	06:37:28	06:43:14	00:05:46
AIM03	RU374	26/10/2022	ARANCIONE	ISTRANA	MONTANARI GIOVANNI	CARTA CARTONE	674	Istrana_CARTA 3	08:24:24	08:33:26	00:09:02
AIM03	RU345	26/10/2022	ARANCIONE	ISTRANA	BOMBARDA MIRCO	CARTA CARTONE	672	Istrana_CARTA 1	08:40:12	08:55:09	00:14:57
AIM03	RU282	26/10/2022	ARANCIONE	ISTRANA	AZZOLIN MASSIMO	CARTA CARTONE	678	Istrana_CARTA 7	08:51:48	09:01:42	00:09:54
AIM03	RU347	26/10/2022	ARANCIONE	ISTRANA	SEMENTATO FAUSTO	CARTA CARTONE	675	Istrana_CARTA 4	09:25:10	09:32:09	00:06:59
AIM03	RU72	26/10/2022	ARANCIONE	ISTRANA	MUSAT IONUT	CARTA CARTONE	676	Istrana_CARTA 5	09:51:49	10:02:15	00:10:26
AIM03	RU184	26/10/2022	ARANCIONE	ISTRANA	MITABLINDA TUDOR	CARTA CARTONE	677	Istrana_CARTA 6	10:01:50	10:09:30	00:07:40
AIM03	RU374	26/10/2022	ARANCIONE	ISTRANA	MONTANARI GIOVANNI	CARTA CARTONE	674	Istrana_CARTA 3	10:22:00	10:31:46	00:09:46
AIM03	RU117	26/10/2022	VERDE	QUINTO DI TREVISO	DE GASPARI DAVIDE	CARTA CARTONE	660	Quinto_CARTA 3	10:32:38	10:39:47	00:07:09

Figura 2.4: Dettaglio permanenze nelle aree di controllo dal 26/10/2022 al 26/10/2022.

È possibile, inoltre, generare un report riguardante gli allarmi riguardanti le fermate per ogni automezzo come mostrato in Figura 2.5. Sono elencati i turni la cui somma delle fermate con motore spento (>10 minuti) e delle fermate con motore acceso sia maggiore di 30 minuti. Sono evidenziate in blu le fermate dove viene spento il quadro dell'automezzo, in rosso le fermate maggiori di 30 minuti e in giallo le fermate sotto i 30 min.

Ogni tipologia di mezzo può avere report dedicati riferiti all'attività di competenza anche se non si percorrono percorsi prestabiliti come nel caso della raccolta porta a porta.

Nel caso delle spazzatrici sono stati impostati i parametri riguardanti i chilometri e tempo di spazzamento eseguiti.

RU326 (Automezzo vasca)		Turno: 26/10 04.37 26/10 12.22		Tempo mezzo fermo:	02:00:35
Missione: Riese Pio X_VPL 4					
	Inizio	Durata	Luogo		
	06.26	00:21:33	VIA CENDROLE (RIESE PIO X)		
	07.10	00:15:00	VIA ASOLANA (RIESE PIO X)		
	08.30	00:13:48	VIA MONTE SANTO (RIESE PIO X)		
	08.49	00:37:19	VIA MONTE GRAPPA (RIESE PIO X)		
	10.39	00:16:55	VIA MONTE SANTO (RIESE PIO X)		
	11.01	00:16:00	QUARTIERE LONGHIN (RIESE PIO X)		

Figura 2.5: Allarmi fermate per il periodo dal 26/10/2022 al 26/10/2022.

Altri dati vengono recepiti da un inserimento manuale degli stessi, un ODS ne è l'esempio. L'operatore, una volta terminato il servizio, compila l'ODS e lo consegna per la rendicontazione.

Successivamente i dati inseriti dall'operatore e presenti nell'ODS vengono trasportati in formato informatico. La procedura di consuntivazione e rendicontazione si distingue su due casi:

- * Ordine di Servizio non modificato;
- * Ordine di Servizio modificato.

Il primo caso rappresenta un ODS riguardante un giro di raccolta effettuato da un operatore. Quest'ultimo, ultimato il servizio, si presta a consegnare il singolo modulo ODS e la rendicontazione avviene in automatico. Il consuntivo avviene mettendo a confronto l'orario di timbratura di inizio e fine servizio dell'operatore in questione.

Nel secondo caso, invece, la rendicontazione e la consuntivazione avviene in maniera manuale. Un ODS modificato consiste in 2 o più moduli di Ordine di Servizio, è il caso in cui ad un operatore venga assegnato più di un giro di raccolta, solitamente effettua un giro e un mezzo giro in due zone diverse.

La rendicontazione e consuntivazione manuale si verifica nei casi in cui vi è un cambio autista, il mezzo richiede l'intervento dell'officina per problemi meccanici o semplicemente il mezzo utilizzato ha delle problematiche che non permettono la prosecuzione del servizio, spazzamento.

Altra tipologia di modulo rendicontato è quello relativo ai disservizi, generato a causa di una segnalazione da parte dell'utenza per un mancato svuotamento. Successivamente è organizzato un giro di controllo per sanare i disservizi oppure per verificare le motivazioni del disservizio.

Il modulo SGQ è anch'esso una fonte di informazioni, potendo segnalare difformità e tracciando il comportamento dell'utenza.

Infine, ogni termine del mese, vengono comparati gli orari di timbratura con gli orari della gestione operativa per verificare le corrette corrispondenze degli orari di servizio degli operatori.

Capitolo 3

Strategia di analisi e raccolta dati

3.1 Definizione del set di dati

Nella fase iniziale della definizione della strategia di analisi si vogliono definire quali sono i dati da osservare.

Dagli strumenti descritti nella Sezione 2.5, LeO, ECOS e DBW, possiamo ricavare delle informazioni a noi utili ai fini dell'analisi.

In particolare, grazie alle loro funzionalità, gli strumenti permettono di estrapolare delle basi dati contenute nel corrispettivo applicativo.

Ogni applicativo fa riferimento a specifici ambiti del servizio PAP: LeO è l'applicativo dove vi sono contenuti i dati relativi alle tracce GPS dei mezzi programmati a svolgere il servizio di raccolta PAP fino a questo momento dalla sua introduzione; ECOS contiene i dati relativi agli ordini di servizio (ODS) che comprendono informazioni relative al servizio che si va ad attuare in un determinato comune, ed inoltre informazioni che provengono dalle pesate registrate dai mezzi e dai semirimorchi; infine DBW contiene dati legati alle utenze come anagrafiche, la tipologia di contratto stipulato tra l'utenza e l'ente che eroga il servizio di raccolta PAP e distinzioni tra utenze condominiali o meno.

Si ricorda che tutti gli applicativi sono in utilizzo da parte di Contarina SpA dall'anno 2010, mentre LeO è uno strumento introdotto nell'anno 2020 e i dati al suo interno diventano consistenti nel periodo temporale più vicino nel momento in cui il sistema di navigazione, a cui è legato l'applicativo, viene adottato da tutti i mezzi che erogano il servizio PAP.

Gli obiettivi delle analisi che andranno svolte, mirano al riconoscimento di criticità all'interno del servizio di raccolta PAP.

In particolare, il piano messo in atto per poter svolgere il servizio è prettamente statico, infatti è stato progettato su analisi precedenti basate su dati che non tengono conto di una probabile dinamicità nel tempo di quest'ultimi.

Le risorse necessarie vengono calcolate come spiegato nella Sezione 2.6.1, in base ad una semplice formula che tiene conto di diversi fattori:

* *Utenze totali;*

- * % esposizione contenitore;
- * Utenze effettive $\rightarrow (utenze\ totali * \% \text{esposizione}) / 100$;
- * Ore del turno;
- * Tempo di trasferimento;
- * Numero di scarichi medi;
- * Tempo di singolo scarico;
- * Pausa
- * Operazioni ordinarie
- * Tempo effettivo netto del turno $\rightarrow Ore\ del\ turno - tempo\ di\ trasferimento - (num.\ scarichi * tempo\ di\ singolo\ scarico) - pausa - operazioni\ ordinarie$;
- * Numero contenitori serviti all'ora $\rightarrow (utenze\ effettive / tempo\ effettivo\ netto)$;
- * Prese effettuate per turno $\rightarrow Tempo\ effettivo\ netto * numero\ contenitori\ serviti\ all'ora$;
- * Risorse necessarie $\rightarrow utenze\ effettive / Prese\ effettuate\ per\ turno$ (RISULTATO)

Questa formula viene ripetuta per ogni comune e per ogni tipologia di materiale che il servizio si presta a raccogliere, individuando quindi un numero di risorse necessarie per effettuare il servizio per un determinato rifiuto in un determinato comune.

Le analisi verteranno sui parametri che vengono messi in gioco per il calcolo delle risorse, evidenziando possibili correlazioni tra di essi e mostrando quali di queste variabili possono essere governate e non, al fine di evidenziare miglie e perseguire il servizio di raccolta nella maniera corretta.

Possiamo distinguere due tipologie di analisi che vanno a concentrarsi su due periodi temporali di grandezza differente.

Vogliamo effettuare un'analisi su base dati di un periodo temporale molto recente di un anno, e un'analisi su base dati riguardante un periodo temporale di 5 anni.

3.2 Raccolta dei dati

I dati estrapolati dalle nostri fonti si presentano, alla consegna, in formato di cartella Excel. Excel ci ha permesso di manipolare facilmente i dati e di creare tabelle e grafici per visualizzare i risultati.

Tuttavia, per la fase successiva del progetto che comprende la raccolta e l'analisi dei dati, si utilizzerà il linguaggio R.

R è un linguaggio di programmazione open-source utilizzato comunemente per l'analisi statistica e la visualizzazione dei dati. Ci permetterà di effettuare analisi più complesse ed inoltre di effettuare operazioni di pulizia e raggruppamento.

Inoltre, R ci offre una vasta gamma di librerie e pacchetti che possono essere utilizzati per automatizzare i nostri processi di analisi. In sintesi, Excel ci ha permesso di estrarre i dati e di effettuare una prima analisi preliminare, mentre R ci permetterà di effettuare analisi più approfondite [19].

La prima analisi vuole indagare in particolare sui dati alla base della programmazione dei giri di raccolta PAP tra il periodo di Novembre, dunque su base dati proveniente dall'applicativo LeO.

Da quest'ultima possiamo ricavare una base dati di informazioni sulle tracce GPS provenienti dai vari mezzi impiegati nel territorio.

L'estrazione presenta la struttura mostrata in Figura 3.1:

Giri_LeO		
PK	ID	INTEGER
	Ods	VARCHAR
	ServizioDesignato	VARCHAR
	Data	DATE
	CodiceMezzo	VARCHAR
	AnagraficaAutista	VARCHAR
	IdGiro	INTEGER
	Zona	VARCHAR
	Destinazione	VARCHAR
	Materiale	CHAR
	%Complelamento	INTEGER
	OrarioUscita	TIME
	OrarioRientro	TIME
	DurataSoste	TIME
	DurataTurno	TIME
	DurataNormalizzata	TIME
	%AttivitàTurno	INTEGER
	Lunghezza	FLOAT
	VelocitàMedia	INTEGER
	VelocitàMax	INTEGER

Figura 3.1: Struttura della base dati relativa ai Giri LeO

Otteniamo, quindi, una base dati dove ogni record rappresenta un giro di servizio di raccolta PAP nel territorio, specificando la destinazione del servizio come comune, l'anagrafica dell'operatore che svolge il turno, il mezzo utilizzato, il tipo di materiale che ha servito, la data di svolgimento del turno. Le tracce GPS forniscono inoltre ulteriori informazioni su un determinato record, specificando orari di uscita dalla sede e rientro in sede, la lunghezza del percorso effettuato e informazioni riguardanti la velocità. Inoltre fornisce informazioni sulle pause effettuate dall'operatore, in particolare vengono distinte le fermate e le soste a motore spento. Quest'ultime vengono considerati tali se il motore del mezzo risulta fermo per almeno 1 minuto.

Si vuole specificare che il campo *Materiale* può assumere i seguenti valori:

- * V, per la raccolta di rifiuto VPL;
- * C, per la raccolta di rifiuto Carta e Cartone;
- * U, per la raccolta di rifiuto Umido;
- * S, per la raccolta di rifiuto Secco non riciclabile;
- * E, per la raccolta di rifiuto Vegetale;

Inoltre, i campi *%Complelamento* e *%AttivitàTurno*, rispettivamente rappresentano la percentuale di adozione da parte dell'operatore nel seguire le indicazioni fornite dal

navigatore installato a bordo del mezzo e la percentuale di attività lavorativa effettiva, escludendo quindi pause e operazioni ordinarie, nel turno.

È di nostro interesse, successivamente, poter riuscire a ricavare il numero di letture, cioè quanti contenitori sono stati serviti/vuotati, effettuate da ciascun giro programmato. Siamo interessati, in particolare, alle letture relative al rifiuto Secco non riciclabile e Vegetale, al fine di poter analizzare determinati parametri non presenti per le altre tipologie di rifiuto che verranno spiegate successivamente.

Il procedimento di lettura avviene grazie ad un transponder collocato nel contenitore, l'operatore che si presta allo svuotamento del contenitore, prima di prendere in carico il bidone tramite un lettore (chiamato saponetta) di transponder legge il contenitore e registra la presa in carico.

I transponder collocati nei contenitori possono essere di due tipologie, LF o UHF, ed inoltre un contenitore è dotato di uno e un solo transponder.

Si effettua un'ulteriore estrazione dati, riguardante le letture e avente la struttura mostrata in Figura 3.2:

Letture		
PK	ID	INTEGER
	Data	DATE
	Orario	TIME
	CodiceContenitore	VARCHAR
	Materiale	CHAR
	CodiceTransponderLF	VARCHAR
	CodiceTransponderUHF	VARCHAR
	CodiceMezzo	VARCHAR
	Volumetria	INTEGER
	ComuneUtenza	VARCHAR
	ZonaAmministrativa	VARCHAR

Figura 3.2: Struttura base dati relativa alle letture nei 12 mesi

Possiamo notare come ogni record rappresenti una e una sola lettura, effettuata in una determinata data da un determinato mezzo in uno specifico comune per un tipo di materiale. Possiamo, dunque, identificare un giro di raccolta programmato come l'insieme delle seguenti informazioni: *Data*, *Materiale*, *CodiceMezzo*, *ComuneUtenza*. Notiamo come esistano records che riguardano lo stesso giro di raccolta PAP perché appunto riguardano ognuno una lettura effettuata durante quel turno specifico.

Il campo *ZonaAmministrativa* specifica la tipologia di zona urbana in cui rientra il comune dove è stato svolto il servizio e assume valori i quali:

- * Zona urbanisticamente complessa;
- * Zona Standard;
- * Zona centro storico;
- * Zona a bassa densità.

Il campo *Volumetria* rappresenta il volume del contenitore servito/vuotato, è possibile infatti risalire alle dimensioni di quest'ultimo tramite il suo codice univoco *CodiceContenitore*.

È stata effettuata una terza estrazione, per quanto riguarda la prima analisi, che comprende informazioni sui pesi conferiti dai mezzi impiegati nel territorio a svolgere

il servizio di raccolta PAP.

Siamo interessati, in particolare, ad i pesi che ogni giro di raccolta ha raccolto per materiale. L'attenzione viene posta dunque su i pesi conferiti dai mezzi, impiegati nella raccolta, nei semirimorchi e su i pesi conferiti dai mezzi in impianto di destino. I mezzi, essendo dotati di una vasca dove contenere il rifiuto raccolto prima di scaricarlo o in semirimorchio o in impianto di destino, sono dotati di una determinata capacità in kg e di conseguenza per ultimare un giro di servizio di raccolta PAP necessitano di effettuare almeno uno svuotamento durante quest'ultimo.

Vi sono quindi due sorgenti differenti dalle quali prelevare i dati necessari. La prima estrazione riguarderà dati provenienti dai sistemi di pesatura collocati nei semirimorchi; infatti, un mezzo che decide di scaricare il rifiuto raccolto presso uno di essi necessita di una registrazione per permettere la rendicontazione finale. La seconda estrazione verte, invece, sui dati proveniente dal sistema di pesatura presente in sede, che permette di ottenere il peso del mezzo e del rifiuto al suo interno nel caso l'ultimo scarico venga effettuato in impianto di destinazione.

Le estrazioni presentano le strutture mostrate nelle Figure 3.3, 3.4:

PesateMezziSuSR		
PK	ID	INTEGER
	DataPesata	DATE
	OrarioPesata	TIME
	CodiceSR	VARCHAR
	CodiceMezzo	VARCHAR
	TargaSR	VARCHAR
	TargaMezzo	VARCHAR
	PesoConferito	INTEGER
	RifiutoConferito	VARCHAR

Figura 3.3: Struttura base dati relativa alle pesate sui semirimorchi

PesateMezziImpianto		
PK	ID	INTEGER
	DataRegistrazione	DATE
	SiglaMezzo	VARCHAR
	TargaMezzo	VARCHAR
	RifiutoConferito	VARCHAR
	TaraComplessivaMezzo	INTEGER
	PesoConferitoRifiuto	INTEGER
	OrarioIngressoImpianto	TIME

Figura 3.4: Struttura base dati relativa alle pesate all'impianto di destinazione

Dall'estrazione relativa alle rilevazioni di peso sui semirimorchi, possiamo distinguere ogni scarico su un determinato semirimorchio grazie al campo *DataPesata* e *OrarioPesata*, potendolo associare ad un determinato mezzo a vasca.

Ogni record, identifica quindi un determinato scarico su semirimorchio da parte di un mezzo impiegato nella raccolta PAP, concerne di peso e tipologia di materiale conferito. La seconda estrazione, simile alla prima, si differenzia in quanto ogni record rappresenti le informazioni rilevate ogni qualvolta un mezzo dedito alla raccolta PAP rientra in impianto. Infatti, ogni mezzo in rientro dal servizio viene pesato all'ingresso della sede tramite dei sistemi di pesatura registrando la targa, il codice del mezzo e il peso

complessivo del mezzo a vasca piena.

La seconda analisi vuole indagare, invece, su informazioni riguardanti un periodo temporale più lungo. Quest'ultimo si vede compreso tra l'anno 2017 e 2022, per un totale di 5 anni.

È stata effettuata un'estrazione riguardante le letture/numero di prese effettuate dai mezzi nei giri di raccolta del servizio PAP per i soli rifiuti Secco e Vegetale, nel periodo di tempo da Novembre 2017 ad Ottobre 2022.

Si è voluto analizzare i dati dei rifiuti Secco e Vegetale perché sono le due tipologie che nel servizio PAP hanno un determinato numero di vuotamenti che un'utenza può sfruttare (dettato dal numero di componenti familiari +1).

Solo nell'ultimo anno si hanno a disposizione i dati provenienti dalle tracce GPS dei mezzi, quindi si è deciso di effettuare delle estrazioni fini all'analisi riguardanti informazioni su letture e pesi registrati.

La prima estrazione è relativa alle letture effettuate dagli operatori nel corso del periodo temporale dei 5 anni. Presenta la struttura mostrata in Figura 3.5:

Letture5Anni		
PK	IDLettura	INTEGER
	Data	DATE
	Zona	VARCHAR
	Comune	VARCHAR
	Materiale	VARCHAR
	CodiceOperatore	VARCHAR
	CodiceMezzo	VARCHAR
	NLetture	INTEGER
	Litri	INTEGER

Figura 3.5: Struttura base dati relativa alle letture di 5 anni

Ogni record di tale estrazione rappresenta le letture effettuate in un determinato giro di raccolta, di quale si conoscono il mezzo, l'autista e il materiale raccolto. Un'ulteriore informazione utile è il campo *Litri* infatti esso specifica la somma dei volumi (in litri) serviti tramite le letture eseguite in un corrispondente record.

La seconda estrazione riguarda i dati relativi alla produzione del rifiuto, quindi dei pesi registrati dai sistemi. Presenta la struttura mostrata in Figura 3.6:

Pesi5Anni		
PK	IdPesata	INTEGER
	Mese	DATE
	Materiale	VARCHAR
	Comune	VARCHAR
	Zona	VARCHAR
	KgRaccolti	INTEGER

Figura 3.6: Struttura base dati relativa ai pesi di 5 anni

Notiamo come l'estrazione sia raggruppata per mese, quindi il dato a nostra disposizione è conseguenza di rilevazioni mensili.

Nello specifico il campo *Mese* viene rappresentato in forma numerica, ogni mese è rappresentato dalla data di inizio mese in cifre (01/01/20xx per gennaio, 01/02/20xx per febbraio e così via...).

Ogni record, rappresenta dunque il peso complessivo raccolto in un determinato mese e in un determinato comune per uno specifico materiale.

3.3 Definizione delle basi dati

Per quanto riguarda le analisi, si vuole poter unire tra loro le basi dati in modo da ottenere una base dati solida per ogni analisi che si andrà ad effettuare.

In particolare per l'analisi dei dati riguardanti il periodo temporale di 12 mesi, si vuole poter associare ad ogni record presente nella base dati *Giri_Leo* uno o più record presenti nelle basi dati *Letture*, *PesateMezziSuSR* e *PesateMezziImpianto*.

L'obiettivo è quello di riuscire ad ottenere una base dati dove ogni record rappresenti un giro di raccolta programmato associato ad un relativo numero di letture effettuate e relativo peso, in kg, raccolto.

A tal fine, sono stati individuati i campi necessari per effettuare l'associazione del numero di letture ai corrispettivi giri di raccolta presenti nella base dati *Giri_Leo*:

$$\begin{array}{c} \textit{Giri_Leo}(\textit{Data}, \textit{Destinazione}, \textit{Mezzo}, \textit{Materiale}) \\ \langle - \rangle \\ \textit{Letture}(\textit{Data}, \textit{ComuneUtenza}, \textit{CodiceMezzo}, \textit{Materiale}) \end{array}$$

Dato che, nel caso in cui vi fosse una corrispondenza per un determinato giro, avremmo solo le informazioni del giro stesso, si è deciso di effettuare un'operazione di tipo *Left-Join* in modo che l'associazione formi una tabella dove ogni riga corrisponde ad un giro di raccolta e le sue informazioni, ed abbia associato un numero che rappresenti il conteggio di quante associazioni sono state effettuate con l'operazione.

Si ricorda che la base dati *Letture* è formata da record che rappresentano una lettura ciascuno, quindi vi saranno più record che verranno associati ad un eventuale giro di raccolta della base dati *Giri_Leo*.

Si effettua un'associazione ad ogni record presente nella base dati *Giri_Leo* con i pesi relativi ai mezzi. Avendo a disposizione due basi dati relative alle pesate dei mezzi impiegati nel servizio di raccolta PAP, e ricordando che non sempre un mezzo effettua lo scarico in impianto di destinazione dopo aver effettuato il giro, si procede unendo alla base dati ottenuta in precedenza *Giri_Leo* e *Letture* dapprima la base dati relativa alle pesate sui semirimorchi e successivamente la base dati relativa alle pesate in sede.

Per quanto riguarda la prima associazione, utilizziamo l'operazione di *Left-Join* tra le basi dati *Giri_Leo* e *Letture* e *PesateMezziSuSR*.

L'operazione di *Left-Join* estende tutte le righe della prima tabella con i valori della seconda tabella che soddisfano una determinata condizione di join, mantenendo le righe della prima tabella.

Sono stati individuati i seguenti campi per effettuare l'operazione:

$$\begin{array}{c}
\text{Giri_Leo}\&\text{Letture}(\text{Data}, \text{Mezzo}, \text{Materiale}) \\
\langle - \rangle \\
\text{PesateMezziSuSR}(\text{DataPesata}, \text{CodiceMezzo}, \text{RifiutoConferito})
\end{array}$$

È stata mantenuta la struttura e la visualizzazione dei campi della base dati *Giri_Leo&Letture*, aggiungendo la colonna *PesoConferito* come somma del campo corrispondente.

Si nota che, se un mezzo effettua due o più scarichi, si avrà più di una coincidenza per l'associazione, e a tal fine il campo *PesoConferito* viene calcolato come somma dei pesi conferiti da un determinato mezzo. Otteniamo la base dati corrispondente *Giri_Leo&Letture&PesateSR*.

La stessa operazione viene effettuata per associare le pesate in impianto di destinazione alla base dati ottenuta in precedenza, e si vuole ricordare che non tutti i mezzi effettuano lo scarico in sede quindi vi è un'alta probabilità di non riuscire ad associare qualche giro di raccolta con qualche record della base dati *PesateMezziImpianto*.

Si utilizza l'operazione di *Left-Join*, in modo da ottenere l'informazione nel caso in cui un determinato giro di raccolta non abbia effettuato lo scarico in sede. Risulterà che il peso conferito da un determinato giro è zero.

È stata mantenuta la struttura e la visualizzazione dei campi della base dati *Giri_Leo&Letture&PesateSR*, aggiungendo la colonna *PesoImpianto* come somma del campo corrispondente.

Successivamente per associare ad ogni giro un solo peso di conferimento, in modo da visualizzarlo come quantità di kg gestita da un determinato servizio di raccolta programmato, ci si presta a sommare i campi:

$$\text{PesoConferito} + \text{PesoImpianto} = \text{PesoTotale}$$

Si vuole aggiungere un campo personalizzato, *Settimana*, visualizzato nel seguente formato: yyyy-ww, dove verrà visualizzato l'anno e il numero della settimana dell'anno corrispondenti alla data a cui fa riferimento una specifica riga. Tale campo viene inserito in quanto si hanno a disposizione più di un giro per un determinato servizio di raccolta PAP per un determinato materiale in una settimana.

Per quanto riguarda l'analisi prevista nel periodo temporale più lungo, si vuole associare ad ogni lettura estratta i relativi pesi raccolti.

A tal punto, si è deciso di effettuare un raggruppamento mensile sulla base *Letture5Anni* mostrata in Figura 3.5 in modo da poter associare le relative informazioni dei pesi mensili a dati anch'essi mensili, ottenendo una base dati equivalente, avendo come numero di letture per ogni riga il campo *NLetture* cioè una somma di tutte le letture effettuate in quel determinato mese.

Individuiamo i seguenti campi per effettuare l'associazione:

$$\begin{array}{c}
\text{Letture5Anni}(\text{Mese}, \text{Comune}, \text{Materiale}) \\
\langle - \rangle \\
\text{Pesi5Anni}(\text{Mese}, \text{Materiale}, \text{Comune})
\end{array}$$

e otteniamo una base dati avente la struttura mostrata in Figura 3.7.

Il campo *Letture* rappresenta la somma delle letture nel mese, il campo *Litri* rappresenta la somma di litri associata alle letture nel mese, il campo *Kg* rappresenta

Letture5Anni&Pesi5Anni	
Id	INTEGER
Mese	DATE
Materiale	VARCHAR
Comune	VARCHAR
Zona	VARCHAR
Utenze	INTEGER
Letture	INTEGER
Litri	INTEGER
%esposizione	INTEGER
Kg	INTEGER
Densità	FLOAT

Figura 3.7: Struttura base dati relativa alle letture associate ai relativi pesi nei 5 anni.

il dato dei pesi mensili associato al mese delle letture effettuate presente nella base dati *Pesi5Anni*, il campo *Utenze* rappresenta il numero di contenitori potenziali che possono essere esposti in quel comune per quel rifiuto, il campo *%esposizione* rappresenta la percentuale di esposizione dei contenitori rispetto al totale cioè al campo *Utenze*, ed infine il campo *Densità*, campo calcolato tramite la formula $Kg / Litri$, che rappresenta la densità media del rifiuto presente nei contenitori, cioè quanto pieno effettivamente è un bidone che ci si presta a servire in un determinato mese dell'anno.

Capitolo 4

Analisi critica dei dati

4.1 Studio preliminare dei dati

Nel presente capitolo, verrà eseguito uno studio premilinare dei dati di due serie storiche utilizzando il linguaggio R. L'obiettivo di questo studio è quello di esplorare i dati e acquisire una comprensione di base dei dati prima di procedere con analisi più avanzate.

Le basi che si andranno ad esplorare e comprendere più a fondo sono rispettivamente le serie riguardanti il periodo di 12 mesi e le serie riguardanti il periodo di 5 anni. In entrambe le categorie, individuiamo delle serie storiche riguardanti alcuni parametri (come le letture, i litri, la % di esposizione, i kg e la densità) che si sviluppano in periodi temporali diversi.

Si andrà ad analizzare in primis la base dati nel periodo temporale di 12 mesi, e successivamente la base dati nel periodo temporale più lungo.

In particolare, l'analisi della serie storica di 12 mesi sarà utilizzata per individuare tendenze a breve termine, comparando quest'ultime tra di loro per poter evidenziare la presenza di eventuali stagionalità nel periodo di un anno di turni di raccolta.

L'analisi riguardante il periodo a lungo termine di 5 anni sarà utilizzata per individuare componenti quali stagionalità, tendenza e ciclicità.

Essendo quest'ultima serie storica una fonte di maggiori quantità di dati rispetto alla precedente, si aggiunge un'analisi per verificare la possibilità di effettuare una previsione di valori come le letture.

4.2 Metodologia

Nell'affrontare l'analisi di serie storiche sfruttiamo le potenzialità del linguaggio R e dell'IDE RStudio [25], open-source e disponibile per più sistemi operativi. Il linguaggio è stato creato e sviluppato specificamente per l'analisi dei dati, rendendolo uno standard in ambito statistico.

Rispetto a Microsoft Excel, R, sebbene meno intuitivo, ha la capacità di riprodurre analisi con set di dati molto diversi.

Un'altra differenza risiede nell'esecuzione rapida di funzioni statistiche e aritmetiche, Excel potrebbe essere la scelta migliore, poiché fornisce una GUI semplice ed intuitiva.

D'altro canto, R può eseguire un'enorme quantità di analisi, è ottimo per identificare tendenze e può decidere quanto siano affidabili tali statistiche. R consente di pulire e organizzare i dati, offrendo più opzioni di visualizzazione.

Esso è dotato di molti comandi e funzioni volti allo studio di quest'ultime. Offre la possibilità di leggere e importare una vasta varietà di dati e in diversi formati.

Dispone di pacchetti preinstallati (packages) come *stats*, con la possibilità di utilizzarne ulteriori al fine delle analisi come ad esempio *tseries*.

Un altro punto forte del linguaggio R è la community che sostiene il progetto, infatti questa è fondamentale per il ciclo di vita del progetto perché permette di sostenerlo offrendo la condivisione di pacchetti, librerie e forum di utenti.

4.3 Analisi Critica

4.3.1 Serie storiche

Una serie storica è un insieme finito di osservazioni, ordinate nel tempo, volte ad esprimere la dinamica di un certo fenomeno.

"Indicando con Y il fenomeno ed Y_t una osservazione all'istante di tempo t , con t numero intero variabile da 1 a N , possiamo esprimere il concetto di serie storica come: $Y_t = Y_1, Y_2, Y_3, \dots, Y_N$ con lunghezza N " [26].

Ogni osservazione relativa al fenomeno che si evolve nel tempo è associabile ad una variabile aleatoria X_t , facente parte di un determinata famiglia di variabili.

L'inferenza sulla serie storica è una parte fondamentale, in quanto presuppone che ogni osservazione derivi da altrettante variabili aleatorie dipendenti e quindi possa riportare alla configurazione del processo generatore della serie stesse.

Le serie storiche si possono suddividere in due categorie [26]:

- * Serie storica di tipo *deterministico*, dove ogni osservazione del fenomeno interessato può essere esattamente determinata da osservazioni precedenti;
- * Serie storica di tipo *stocastico*, dove ogni osservazione del fenomeno interessato può essere determinata da alcuni valori parziali delle osservazioni precedenti.

Esistono due approcci per l'analisi delle serie storiche come citato in [3].

L'approccio tradizionale assume che il processo alla base della serie storica sia composto da una parte deterministica, che permette la sua scomposizione in componenti quali di tendenza, stagionali e/o ciclici, e che la differenza tra le osservazioni effettuate e i dati teorici sia attribuibile ad una componente casuale di residuo. L'approccio moderno invece, assume che il processo alla base della serie storica sia frutto di un processo stocastico descrivibile tramite un modello probabilistico di tipo parametrico.

4.4 Approccio tradizionale all'analisi

L'approccio tradizionale all'analisi di serie storiche è conosciuto come "Box-Jenkins" o "metodo ARIMA" (Autoregressive Integrated Moving Average) [17].

Si assume che i dati in una serie storica relativi ad un fenomeno Y siano generati da un processo del tipo:

$$Y_t = f(t) + u_t$$

dove $f(t)$ genera una sequenza completamente deterministica, e u_t è una sequenza di variabili aleatorie che obbediscono ad una qualche legge di probabilità (parte stocastica della serie).

Nell'approccio tradizionale si assume che $f(t)$ sia l'evoluzione temporale del fenomeno, e che i residui (cioè la differenza residua tra i valori teorici e i valori osservati) siano associabili ad errori accidentali e quindi dovuti al caso.

I residui vengono indicati con ϵ_t e sono intesi come variabili aleatorie indipendenti (cioè il verificarsi di un evento non influisce sulla probabilità del verificarsi di un altro), identicamente distribuite (distribuzione della probabilità di una variabile in base a qualche fenomeno aleatorio), con media nulla e varianza costante.

La parte sistematica, nell'approccio moderno, viene considerata come ipoteticamente mancante o già eliminata mediante stime o altri modelli e viene studiata la componente stocastica, cioè la serie senza i disturbi.

La parte sistematica è la risultante di tre componenti che non sono osservabili in modo diretto:

- * *Trend*: componente di tendenza di fondo del fenomeno, è il comportamento di lungo periodo che rappresenta l'evoluzione strutturale del fenomeno osservato e ci permette di evidenziare un andamento crescente/decrecente;
- * *Ciclo*: componente evidenziata da un'alternanza di fluttuazioni di segno diverso intorno al trend, si presenta sottoforma di contrazioni o espansioni del fenomeno in esame;
- * *Stagionalità*: componente costituita da variazioni che si riscontrano con analoga intensità negli stessi periodi di anno in anno, con intensità diversa nel corso dell'anno, determinando cicli di periodi.

Una serie storica può essere descritta da ulteriori diverse proprietà statistiche, come ad esempio la media, la deviazione standard, in aggiunta alle componenti descritte in precedenza.

La componente *accidentale* o di *disturbo* o *casuale* è data dalle variazioni residue le quali non possono essere caratterizzate dalle componenti precedenti.

Le componenti appartenenti alla parte sistematica possono essere individuate da metodi di scomposizione. In particolare essi tendono ad individuare i pattern di stagionalità e di trend-ciclo. Quest'ultimo pattern è una combinazione dei due pattern trend e ciclo perché il trend osservato in un lungo periodo quindi con un intervallo di osservazione molto ampio può rilevarsi come parte di un ciclo.

Una serie storica può essere descritta mediante delle tipologie di modelli tradizionali:

- * Modello additivo: $Y_t = T_t + C_t + S_t + \epsilon_t$;
- * Modello moltiplicativo: $Y_t = T_t \cdot C_t \cdot S_t \cdot \epsilon_t$.

dove T_t corrisponde al valore della componente *trend* nel tempo t , C_t corrisponde al valore della componente ciclo, S_t corrisponde al valore della componente stagionalità e ϵ_t corrisponde al valore della componente accidentale.

Il modello additivo presuppone che le componenti che lo hanno generato siano tra di loro indipendenti, mentre il modello moltiplicativo presuppone che le componenti siano tra di loro interdipendenti. Il modello moltiplicativo può essere trasformato in additivo usando una somma dei logaritmi delle componenti del modello additivo:

$$\log y_t = \log TC_t + \log S_t + \log U_t.$$

In conclusione, nell'approccio tradizionale le componenti sono rappresentate secondo canoni matematici, infatti il trend viene definito come una costante o funzione polinomiale, la stagionalità come una sommatoria di valori che si ripete in periodi in modo regolare.

L'unica componente di natura stocastica è la componente che rappresenta le irregolarità, cioè la componente dei residui, che solitamente viene rappresentata come un *white-noise*, togliendo la sua correlazione con gli errori passati.

L'analisi delle serie storiche ad approccio tradizionale, nel tempo, si è evoluta verso un approccio più moderno, dove la serie storica in considerazione viene valutata secondo l'ipotesi che la parte sistematica venga a mancare focalizzando l'attenzione dell'analisi sulla componente stocastica.

Uno degli scopi fondamentali dell'analisi tradizionale delle serie temporali è quello di scomporre la serie nelle sue componenti, isolandole per poterle studiare meglio.

4.5 Approccio moderno all'analisi

L'obiettivo dell'approccio è quello di usare la serie storica osservata per individuare il processo stocastico stazionario che si adatti maggiormente ai dati reali, da utilizzare per calcolare previsioni future della serie stessa [1].

I processi stocastici stazionari godono di particolari proprietà probabilistiche che risultano utili per modellare molte serie storiche che si incontrano nella realtà.

I due approcci, pur ponendo quesiti diversi, sono matematicamente equivalenti, come dimostrato da Wiener e Khintchine, essendo semplicemente metafore diverse per concepire e trattare i dati [27].

Ci serviamo quindi, della nozione di processo stocastico, modello utile per descrivere e modellare il comportamento probabilistico di un certo fenomeno che può evolvere nel tempo [21].

Un processo stocastico X_t (chiamato anche processo aleatorio) è una famiglia di variabili casuali descritte da un parametro t appartenente ad un insieme parametrico T , detto insieme dei tempi.

Una serie storica è un insieme finito di osservazioni ordinate nel tempo e solitamente misurate in intervalli di tempo equispaziati (settimanale, mensile, annuale...).

Possiamo definire quindi una serie storica come una realizzazione di un insieme di variabili aleatorie X_t , e quindi, può essere vista come una realizzazione finita di un

processo stocastico.

L'analisi di una serie storica può avere molteplici obiettivi, tra cui la descrizione del fenomeno osservato nel tempo evidenziando sia eventuali irregolarità che regolarità; l'individuazione e la spiegazione del processo generatore del fenomeno osservato ed infine la previsione dell'andamento del fenomeno nel futuro.

4.6 Analisi dati 12 mesi

Il primo passo nell'analisi dei dati riguardanti il periodo temporale di 12 mesi è quello di comprendere al meglio l'andamento di vari parametri in un periodo di tempo compreso tra Novembre 2021 e Ottobre 2022.

In questo studio, verrà effettuata un'analisi storica dei dati prelevati in 12 mesi al fine di individuare possibili tendenze preliminari di determinati parametri e per la comprensione del comportamento del servizio di raccolta PAP nell'ultimo anno.

Questa analisi rappresenta un passo importante per comprendere il contesto in cui l'azienda opera e per formulare strategie a lungo termine efficaci.

4.6.1 Raccolta dati e definizione dataset

Essendo le operazioni di consegna e raccolta dati preliminari svolte nell'ambiente di software Excel, tutti i dati vengono importati nell'ambiente di lavoro RStudio, tramite l'installazione del pacchetto *readxl* [15].

read_excel() è una funzione del pacchetto *readxl* in R che consente di importare i dati da un file Excel in R. Questa funzione accetta diversi argomenti che controllano il modo in cui i dati vengono importati.

Gli argomenti principali della funzione sono:

- * **path**: il percorso al file Excel da importare;
- * **sheet**: il nome del foglio da importare. Se si desidera importare più fogli, è possibile utilizzare un vettore di stringhe per specificare i nomi dei fogli;
- * **col_names**: un flag che indica se i nomi delle colonne devono essere importati dalla prima riga del foglio o se devono essere assegnati manualmente. Se impostato su TRUE, i nomi delle colonne verranno importati dalla prima riga del foglio. Se impostato su FALSE, verranno assegnati automaticamente i nomi delle colonne V1, V2, V3, etc.;
- * **skip**: il numero di righe da saltare all'inizio del foglio prima di importare i dati. Questo può essere utilizzato per ignorare le righe di intestazione. *na*: una stringa che indica i valori mancanti nei dati. I valori mancanti verranno sostituiti con i valori NA in R;
- * **range**: un intervallo di celle da importare, se non si desidera importare l'intero foglio.

Una volta che i dati sono stati importati in R utilizzando la funzione *read_excel()*, possono essere elaborati e analizzati utilizzando le molte funzioni disponibili in R per

l'elaborazione dei dati.

Si effettua la creazione di due dataset, sui quali verteranno le operazioni di analisi, differenziati per tipologia di materiale raccolto.

Come detto in Sezione 3.2, si ha a disposizione solamente records riguardanti la raccolta di rifiuto Secco non riciclabile e rifiuto Vegetale (vedi Listing 4.1).

Listing 4.1: Creazione dataset Secco e Vegetale in RStudio

```
Secco <- Dataset[Dataset$Materiale=="S",]
Vegetale <- Dataset[Dataset$Materiale=="E",]
```

Si ottengono due dataframe, rispettivamente *Secco* e *Vegetale*, che corrispondono ai dati prelevati in principio dal foglio Excel filtrati per tipologia di materiale raccolto.

L'obiettivo di questa analisi è quello di esplorare il comportamento di determinati parametri nel dataset.

Prima di iniziare l'analisi di ogni parametro, i dati sono stati puliti tramite l'eliminazione degli outliers. Questa pulizia è stata eseguita per garantire che i risultati dell'analisi siano rappresentativi e non influenzati da valori anomali.

4.6.2 Analisi preliminare

Si vogliono analizzare i dati relativi alla durata dei giri di raccolta programmati, alle letture effettuate, al volume in litri raccolto, ai kg di rifiuto trattato e infine i dati relativi alla densità dei contenitori.

Una volta caricati i dati, è importante esplorare i dati e comprendere la loro struttura.

Sfruttiamo il comando `summary()`, una funzione built-in in R che fornisce una sintesi statistica dei dati contenuti in un data frame. Questa funzione produce una tabella, mostrata in Listing 4.2, che riassume alcune informazioni statistiche chiave sui dati, come la media, la deviazione standard, il minimo, il massimo, il primo quartile (Q1), il mediano (Q2) e il terzo quartile (Q3).

Listing 4.2: output funzione `summary()` per ogni parametro in analisi

```
> summary(Vegetale$Letture)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  2.00   70.00   96.00   95.23  120.00  292.00
> summary(Secco$Letture)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  1.0   121.0   143.0   144.5   166.8   554.0
> summary(Vegetale$Volume)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  360   12840   18000   17996   22800   57120
> summary(Secco$Volume)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  120   15668   18600   18819   21960   68720
> summary(Vegetale$`Peso Totale raccolto`)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  50    1520    2190    2374    2980   12630
```

```

> summary(Secco$`Peso Totale raccolto`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    10   1860    2270    2320   2720    9850
> summary(Vegetale$Densita)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.001956 0.109657 0.126843 0.142369 0.147410 8.166667
> summary(Secco$Densita)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.007083 0.111483 0.123712 0.162743 0.137292 24.833333

```

Le statistiche indicate in Listing 4.2 forniscono una panoramica della distribuzione dei dati, e possono essere utilizzate per identificare eventuali tendenze o outlier.

Si nota che se la differenza tra il valore minimo e il valore massimo è molto grande, potrebbe esserci un outlier che influisce sulla distribuzione dei dati e questo si verifica in ogni dataset che si vuole analizzare.

Inoltre media e deviazione standard sono due statistiche importanti che descrivono la concentrazione dei dati intorno alla media.

La media è il valore che rappresenta la posizione centrale dei dati, ed è calcolata come la somma di tutti i valori divisa per il numero di elementi. Essa è un buon indicatore del valore tipico o del valore che si può aspettare di ottenere se si effettua un'osservazione.

La deviazione standard è una misura della dispersione dei dati intorno alla media. Più precisamente, la deviazione standard descrive la quantità di scostamento tra i valori individuali e la media. Se la deviazione standard infatti è elevata, significa che i valori sono dispersi su una vasta gamma di valori e che la media non descrive molto bene la concentrazione dei dati. Se, invece, la deviazione standard è bassa, significa che i valori sono concentrati intorno alla media.

Si sfrutta la funzione `sd()`, come mostrato in Listing 4.3, per calcolare la deviazione standard di una colonna di un data frame.

Listing 4.3: output funzione `sd()` per ogni parametro in analisi

```

> sd(Vegetale$Letture)
[1] 39.24114
> sd(Secco$Letture)
[1] 46.44749
> sd(Vegetale$Volume)
[1] 7521.339
> sd(Secco$Volume)
[1] 6048.346
> sd(Vegetale$`Peso Totale raccolto`)
[1] 1307.177
> sd(Secco$`Peso Totale raccolto`)
[1] 869.0956
> sd(Vegetale$Densita)
[1] 0.1944461
> sd(Secco$Densita)
[1] 0.7590151

```

Utilizzando queste due statistiche insieme ed avendo osservato ogni differenza elevata tra massimo e minimo di ogni parametro, si nota che i dati necessitano di una pulizia per parametro, avendo una determinata media e una determinata deviazione standard elevata.

4.6.3 Pulizia dei dataset

Si effettua la pulizia dei dati per ogni parametro e per ogni tipologia di rifiuto, come mostrato in Listing 4.4.

Listing 4.4: Pulizia dei dati tramite outliers in base ai parametri da analizzare

```
# PULIZIA

# DURATA NORMALIZZATA
outliers <- boxplot.stats(Vegetale$`Durata Normalizzata`)$out
DurataEClean <- Vegetale[-which(Vegetale$`Durata Normalizzata` %
  in% outliers),]

outliers <- boxplot.stats(Secco$`Durata Normalizzata`)$out
DurataSClean <- Secco[-which(Secco$`Durata Normalizzata` %in%
  outliers),]

# LETTURE
outliers <- boxplot.stats(Vegetale$Letture)$out
LettureEClean <- Vegetale[-which(Vegetale$Letture %in% outliers)
  ,]

outliers <- boxplot.stats(Secco$Letture)$out
LettureSClean <- Secco[-which(Secco$Letture %in% outliers),]

# VOLUME
outliers <- boxplot.stats(Vegetale$Volume)$out
VolumeEClean <- Vegetale[-which(Vegetale$Volume %in% outliers),]

outliers <- boxplot.stats(Secco$Volume)$out
VolumeSClean <- Secco[-which(Secco$Volume %in% outliers),]

# PESI
outliers <- boxplot.stats(Vegetale$`Peso Totale raccolto`)$out
KgEClean <- Vegetale[-which(Vegetale$`Peso Totale raccolto` %in%
  outliers),]

outliers <- boxplot.stats(Secco$`Peso Totale raccolto`)$out
KgSClean <- Secco[-which(Secco$`Peso Totale raccolto` %in%
  outliers),]

# DENSITA'
outliers <- boxplot.stats(Vegetale$Volume)$out
DensitaEClean <- Vegetale[-which(Vegetale$Volume %in% outliers),]
outliers <- boxplot.stats(Vegetale$`Peso Totale raccolto`)$out
DensitaEClean <- Vegetale[-which(Vegetale$`Peso Totale raccolto`
  %in% outliers),]
```

```

outliers <- boxplot.stats(Vegetale$Densita)$out
DensitaEClean <- Vegetale[-which(Vegetale$Densita %in% outliers)
,]

outliers <- boxplot.stats(Secco$Volume)$out
DensitaSClean <- Secco[-which(Secco$Volume %in% outliers),]
outliers <- boxplot.stats(Secco$`Peso Totale raccolto`)$out
DensitaSClean <- Secco[-which(Secco$`Peso Totale raccolto` %in%
  outliers),]

outliers <- boxplot.stats(Secco$Densita)$out
DensitaSClean <- Secco[-which(Secco$Densita %in% outliers),]

```

È di nostro interesse, dopo aver effettuato la pulizia dei dati, comprendere la distribuzione di essi. In R, la funzione `hist()` può essere utilizzata per creare un istogramma dei valori di una serie.

L'istogramma è una rappresentazione grafica della distribuzione dei valori di una serie, mostra la frequenza con cui si verificano valori all'interno di intervalli di valori specifici. In Figura 4.1 vengono mostrati gli istogrammi delle letture.

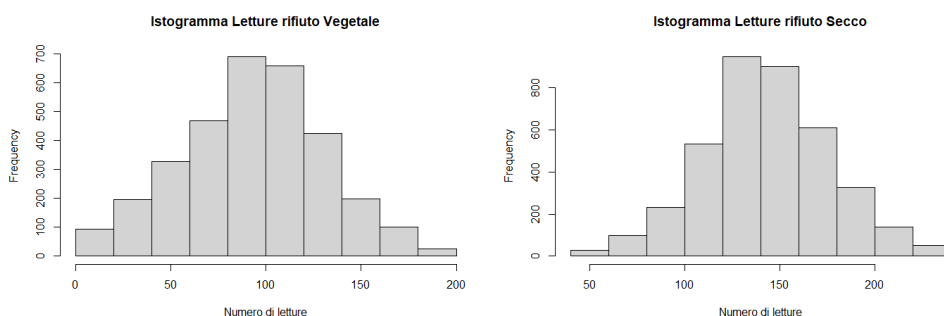


Figura 4.1: Istogrammi relativi alle letture effettuate dai giri di raccolta rispettivamente per rifiuto Vegetale e Secco.

Per quanto riguarda le letture, si può evidenziare in entrambi gli istogrammi di Figura 4.1 una distribuzione a campana simmetrica, dove la maggior parte dei valori si concentra intorno alla media e i valori estremi sono più rari. Si individua una distribuzione quindi, normale, che è una distribuzione di probabilità continua che descrive i valori di una variabile casuale distribuita intorno a una media. La distribuzione normale è spesso descritta come la distribuzione più importante in statistica e appare in molte applicazioni pratiche.

Le operazioni di pulizia precedentemente effettuate fanno notare come la deviazione standard si sia abbassata ulteriormente. Infatti, più piccola essa è, e più i valori sono concentrati strettamente intorno alla media, mentre se la deviazione standard è grande, i valori sono più diffusi e meno concentrati intorno alla media.

Le conclusioni tratte dagli istogrammi relativi alle letture si possono applicare contestualmente agli istogrammi relativi alla volumetria, in Figura 4.2 e ai kg di rifiuto raccolti, in Figura 4.3.

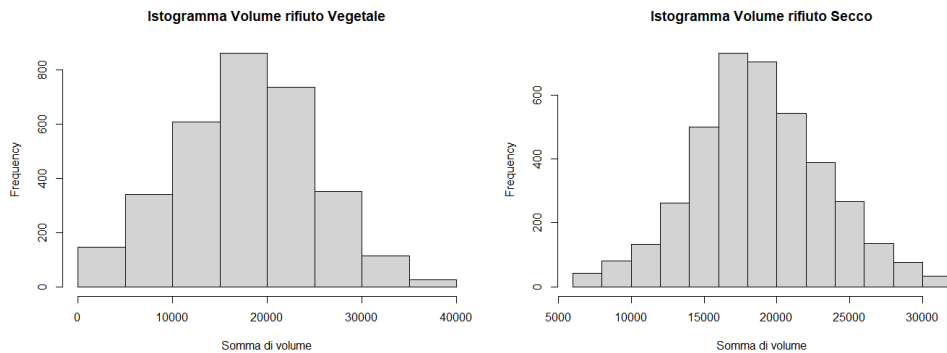


Figura 4.2: Istogrammi relativi alla volumetria raccolta dai giri di programmati rispettivamente per rifiuto Vegetale e Secco.

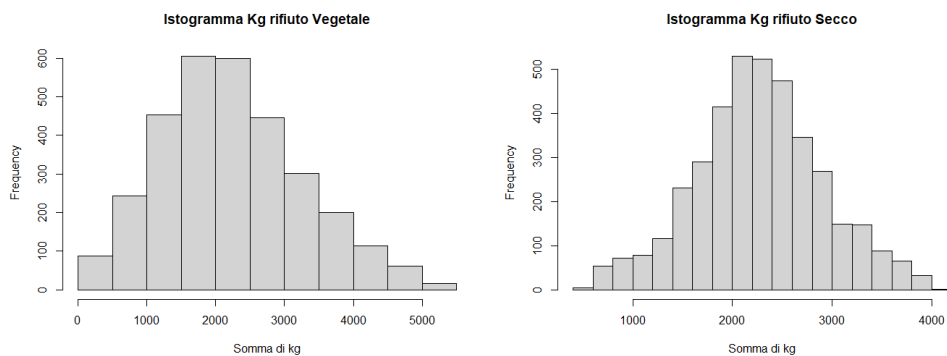


Figura 4.3: Istogrammi relativi ai kg di rifiuto raccolti dai giri di programmati rispettivamente per rifiuto Vegetale e Secco.

In molti casi, la distribuzione normale viene utilizzata per modellare i dati e per effettuare previsioni. Ad esempio, può essere utilizzata per calcolare la probabilità di un evento futuro basato sui dati storici.

Ora si utilizza la funzione `fitdist()` [12] in R.

Quest'ultima è una funzione che appartiene al pacchetto `fitdistrplus` ed è utilizzata per adattare una distribuzione probabilistica a un insieme di dati.

La funzione calcola i parametri della distribuzione che meglio descrivono i dati forniti, inoltre supporta una vasta gamma di distribuzioni, tra cui normali, esponenziali e tante altre.

La funzione `fitdist` richiede in input i dati da modellare e la distribuzione desiderata e ritorna un oggetto di tipo `"fitted.distribution"` che contiene informazioni sui parametri ottimali della distribuzione e altre statistiche relative all'adattamento. Questo oggetto può essere utilizzato per visualizzare la distribuzione ottimale e confrontare la sua aderenza ai dati originali.

La funzione genera inoltre diversi grafici per visualizzare l'aderenza della distribuzione ottimale ai dati originali, tra cui:

* Istogramma dei dati:

L'istogramma dei dati è un grafico a barre che mostra la distribuzione dei valori

presenti nei dati. Ogni barra rappresenta un intervallo di valori e la sua altezza indica il numero di valori presenti in quell'intervallo.

* Curva di densità ottimale:

La curva di densità ottimale è un grafico che mostra la distribuzione probabilistica ottimale che meglio descrive i dati. La curva viene tracciata sull'istogramma dei dati e mostra la probabilità che un valore casuale appartenente ai dati sia compreso in un intervallo specifico. La forma della curva di densità ottimale fornisce informazioni sulle proprietà della distribuzione, come la posizione centrale, la dispersione e la simmetria.

* Q-Q plot:

Il Q-Q plot, o quantile-quantile plot, è un grafico che mostra la relazione tra la distribuzione quantile dei dati e quella della distribuzione ottimale. I quantili dei dati sono i valori che dividono i dati in parti uguali e sono ordinati dal più piccolo al più grande. Il Q-Q plot rappresenta i quantili dei dati sull'asse orizzontale e i quantili della distribuzione ottimale sull'asse verticale. Se la distribuzione ottimale descrive adeguatamente i dati, i punti del Q-Q plot dovrebbero seguire una linea retta.

* Curve di sopra/sotto:

Le curve di sopra e di sotto sono grafici che mostrano l'intervallo di confidenza della distribuzione ottimale. L'intervallo di confidenza indica la quantità di incertezza nei parametri della distribuzione ottimale e indica la probabilità che i veri parametri siano compresi in un intervallo specifico.

Notiamo che dal primo grafico di Figura 4.4, l'istogramma dei dati, mostra la distribuzione dei valori presenti nei dati. Se i dati seguono una distribuzione normale, l'istogramma dovrebbe mostrare una forma approssimativamente simmetrica con un picco centrale e una graduale riduzione dei valori ai bordi.

Dal grafico Q-Q plot si nota che i dati seguono una distribuzione normale, cioè i punti del Q-Q plot dovrebbero seguire una linea retta molto vicina alla diagonale. Questo indica una buona aderenza tra i dati e la distribuzione normale.

Infine, le curve di sopra e di sotto mostrano l'intervallo di confidenza della distribuzione ottimale. I dati, seguendo una distribuzione normale, mostrano come l'intervallo di confidenza sia molto stretto e comprende la forma della distribuzione normale.

In sintesi, i 4 grafici generati da *fitdist()* e mostrati in Figura 4.4 in R sono molto utili per valutare se i dati seguono una distribuzione normale e per visualizzare la precisione dell'adattamento.

I grafici mostrano una forma simile alla distribuzione normale e una buona aderenza tra i dati e la distribuzione normale, ciò indica che i dati possono essere modellati con successo con una distribuzione normale.

Sono stati visualizzati ed analizzati ulteriori grafici per quanto riguarda gli altri parametri che andranno in analisi, i relativi grafici sulla volumetria relativa al rifiuto Secco in Figura 4.5, i kg raccolti di rifiuto Vegetale in Figura 4.6.

Si nota come anch'essi ripercuotono la situazione precedentemente sviluppata sulla letture effettuate per il rifiuto Vegetale.

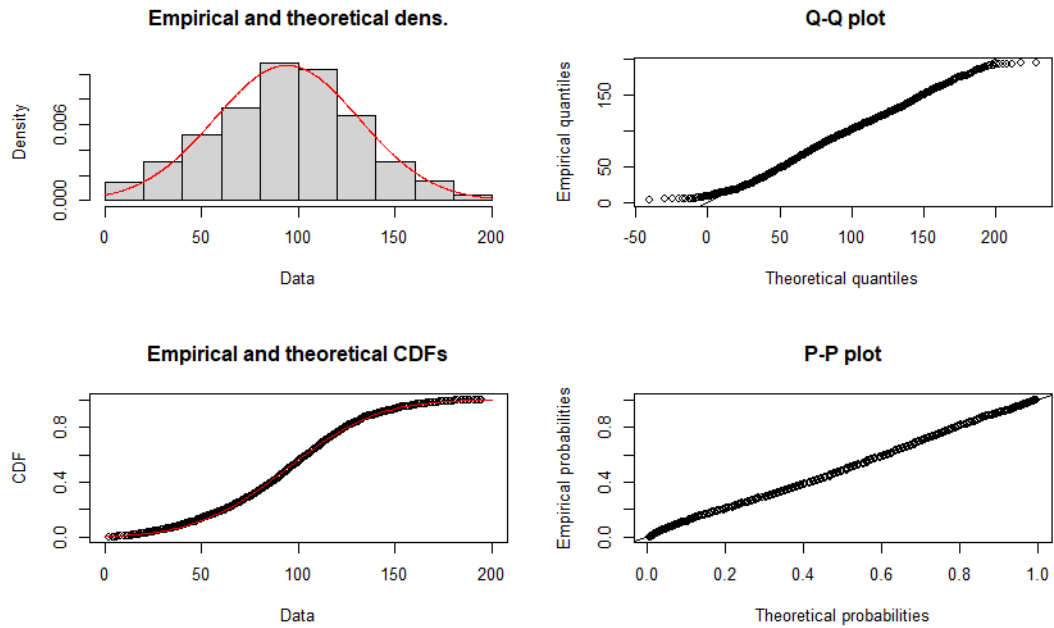


Figura 4.4: Grafici generati dalla funzione *fitdist()* sui dati relativi alle Letture del rifiuto Vegetale.

4.6.4 Analisi

Ora, è di nostro interesse verificare la presenza di correlazione tra le variabili in analisi. Ci sono diversi modi per calcolare se esiste una relazione tra due parametri in R. Uno dei modi più semplici è utilizzare il coefficiente di correlazione.

Il coefficiente di correlazione (Pearson)[16] misura la relazione lineare tra due variabili quantitative. Questo coefficiente può assumere valori tra -1 e 1, dove -1 indica una relazione negativa forte, 1 indica una relazione positiva forte e 0 indica una mancanza di relazione.

Per calcolare il coefficiente di correlazione in R, si può utilizzare la funzione *cor()* [13].

Vengono scelti determinati parametri che possono esser correlati tra loro.

In particolare si vuole evidenziare l'esistenza di correlazione per comprendere al meglio l'andamento nel tempo dei parametri stessi.

La densità di un contenitore, ad esempio, è espressa tramite l'unità di misura Kg/Litri quindi è dipendente sia dal peso totale raccolto per determinati giri di raccolta sia dalla volumetria gestita dai giri che ricordiamo rappresenta la dimensione del contenitore e può esser differente.

Le letture stesse sono un indicatore molto importante nei giri di raccolta, infatti comprendere cosa le porta ad assumere una tendenza in aumento o viceversa può evidenziare una correlazione con altre variabili. Un esempio può essere il caso in cui il numero di letture aumenti ma il contenuto dei contenitori vuotati è relativamente basso.

Queste relazioni aiutano a comprendere l'andamento dei periodi di raccolta rispettiva-

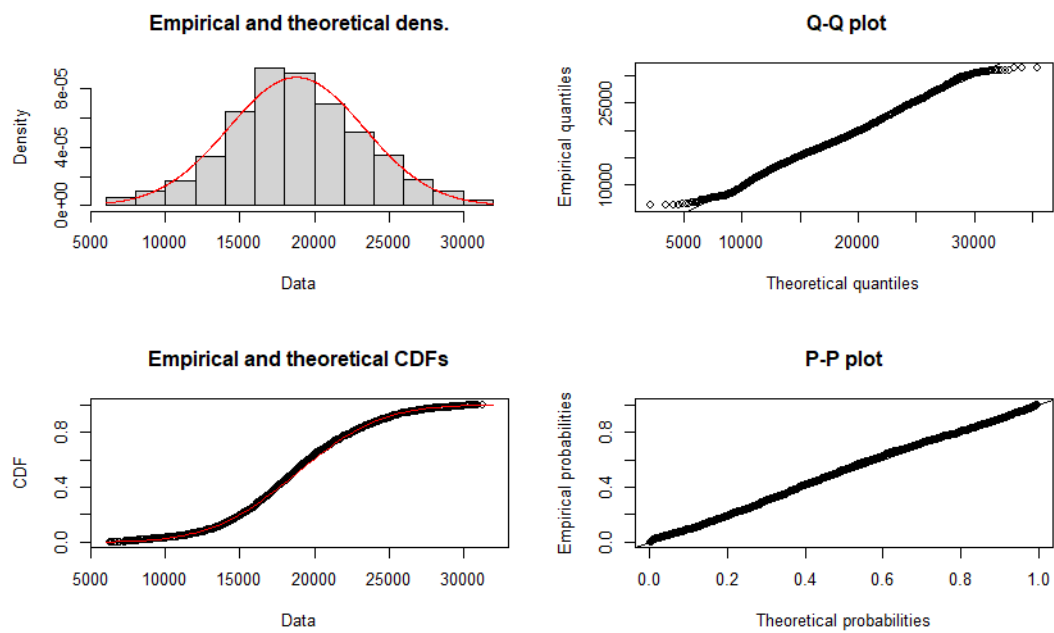


Figura 4.5: Grafici generati dalla funzione *fitdist()* sui dati relativi alla volumetria del rifiuto Secco.

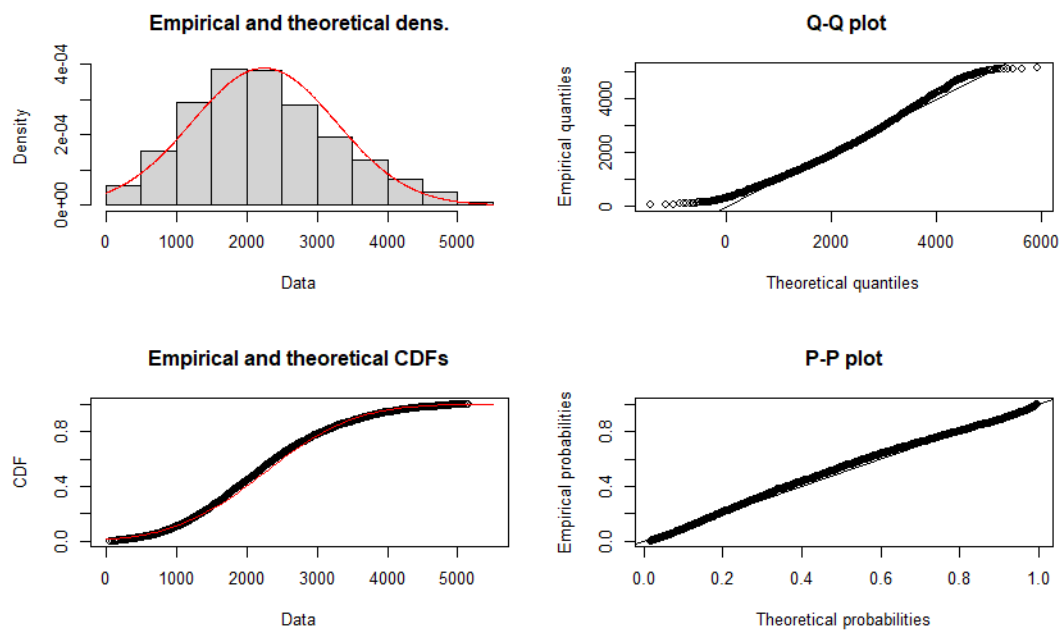


Figura 4.6: Grafici generati dalla funzione *fitdist()* sui dati relativi ai kg raccolti di rifiuto Vegetale.

mente per il rifiuto Vegetale e Secco non riciclabile, potendo verificare la presenza di trend stagionali e a cosa sono dovuti.

Prima di effettuare le operazioni di correlazione, si vuole raggruppare per settimana le rilevazioni, in particolare si vuole visualizzare le letture, i kg come somma di essi, mentre la densità e la volumetria come media settimanale. Otteniamo i valori di output per la funzione `cor()` come mostrato in Listing 4.5.

Listing 4.5: Tabella riassuntiva dei valori di output per rilevazione correlazione tra variabili.

```
> c1<- cor(LettureECleanAcc$Letture_Sum, DensitaECleanAcc$Densita
_mean)
> c2<- cor(LettureSCleanAcc$Letture_Sum, DensitaSCleanAcc$Densita
\ _mean)
> c3<- cor(LettureECleanAcc$Letture_Sum, VolumeECleanAcc$Volume\
_mean)
> c4<- cor(LettureSCleanAcc$Letture_Sum, VolumeSCleanAcc$Volume_
_mean)
>
```

	Letture - Media Densita Volumetria	Letture - Media
Vegetale	0.1479401	0.7106815
Secco	-0.3106624	0.2284629

Il coefficiente di Pearson è una misura statistica della correlazione tra due variabili quantitative. La sua grandezza indica la forza e la direzione della relazione lineare tra le due variabili [16].

Benchè alcuni risultati non siano solidi, possiamo identificare un carattere per le seguenti correlazioni.

Un coefficiente di Pearson vicino a -1 indica una relazione negativa forte. Questo significa che le due variabili sono inversamente correlate, ovvero quando una variabile aumenta l'altra tende a diminuire e viceversa, nel nostro caso le letture in relazione alla media di densità settimanale per il rifiuto Secco.

Un coefficiente di Pearson vicino a 1 indica una relazione positiva forte. Questo significa che le due variabili sono direttamente correlate, ovvero quando una variabile aumenta anche l'altra tende ad aumentare e viceversa, nel nostro caso corrisponde alla relazione che corre tra le letture e la media di volumetria per il rifiuto Vegetale.

Un coefficiente di Pearson vicino a 0 indica una relazione debole o assente. Questo significa che non esiste una relazione lineare significativa tra le due variabili e che altri metodi di analisi potrebbero essere necessari per valutare la relazione tra le due variabili.

L'interpretazione del coefficiente di Pearson dipende dal contesto e dalla scala dei dati e non esiste un valore specifico che possa essere definito come "buono" o "cattivo", ma in linea di massima si può comprendere una tendenza di essi.

Volendo visualizzare graficamente questi dati, si esegue un plot per avere un riscontro grafico sul rifiuto Secco.

Possiamo notare come il grafico a linee in Figura 4.7, che mette in relazione il

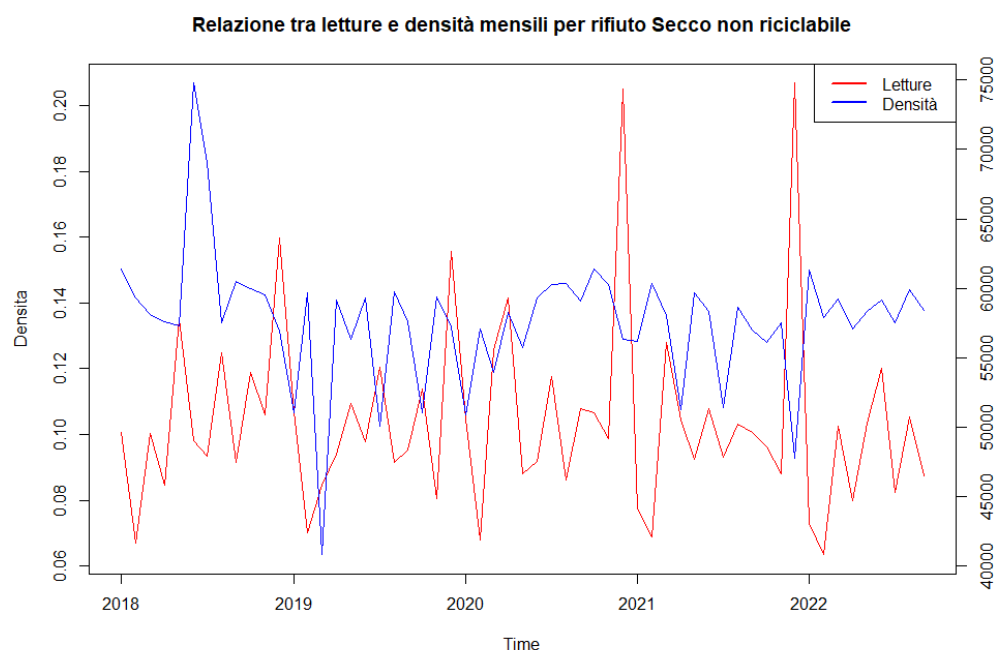


Figura 4.7: Grafico che mette in relazione letture e media di densità per il rifiuto Secco.

numero di letture e la media di densità del rifiuto Secco, è caratterizzato da picchi alterni per entrambe le variabili. Si può notare come le letture del rifiuto Secco non riciclabile aumentino e allo stesso tempo cali la media di densità settimanale in frequenza abbastanza ripetuta.

Questo fenomeno si può spiegare: essendo il rifiuto Secco non riciclabile appartenente ad una categoria di rifiuti che richiedono un pagamento ad ogni vuotamento, le utenze decidono di vuotarlo in determinati periodi anche se il bidone ha una densità di media bassa (cioè nel bidone è contenuto poco rifiuto). Si nota in particolare evidenza il caso di un mese estivo e il mese di dicembre dove entrambi vedono un aumento delle letture ma una densità media in calo. I picchi estivi possono coincidere con un mese dove le utenze non sono nella residenza in quel periodo e quindi decidono di esporre il bidone preventivamente. Mentre nel mese di dicembre avviene il maggior numero di letture perchè la maggior parte delle utenze non ha raggiunto il numero di vuotamenti pagati ed espone il bidone anche se quest'ultimo risulta quasi vuoto.

L'aumentare delle letture effettuate e allo stesso tempo il loro calare, è direttamente proporzionale all'aumentare e al corrispettivo calare della percentuale di esposizione dei bidoni.

Si può quindi concludere affermando che le letture e quindi la percentuale di esposizione dei contenitori è legata alla densità degli stessi.

4.7 Analisi dati 5 anni

La presente sezione ha come obiettivo l'analisi di una serie storica di cinque anni, che va dal 2018 al 2022. La serie in questione verrà decomposta al fine di analizzare le sue componenti, come la stagionalità, il trend e altre eventuali componenti. Questa analisi sarà condotta utilizzando tecniche di decomposizione di serie storiche, che ci permetteranno di identificare le componenti sottostanti che influenzano il comportamento della serie e di capire come queste componenti interagiscono tra loro.

L'obiettivo finale dell'analisi è quello di fornire una comprensione più profonda della serie storica, che sarà utile per prevedere il suo comportamento futuro e supportare decisioni informate.

Si nota che nelle estrazioni riguardanti le letture e i pesi nel periodo di 5 anni, si differenziano per unità di misura temporale di ciascun record.

Infatti le rilevazioni effettuate nella base dati *Pesi5Anni* è di frequenza mensile, mentre la base dati *Lecture5Anni* presenta una frequenza relativa giornaliera e variabile.

4.7.1 Pulizia dei dati

Possiamo formare delle basi dati differenti per degli scopi diversi, si può notare come nella base dati *Lecture5Anni* per alcune date, nei comuni, vi siano dei records aventi un numero di letture molto inferiori e discostanti dagli altri records.

Approfondendo, le righe dove sono presenti numeri di letture bassi sono dei servizi cosiddetti di "recupero", mentre nelle giornate di effettiva raccolta prefissata il numero di letture tende ad esser un numero sostanzialmente appagabile.

Differenziamo le letture di rifiuto Secco non riciclabile e Vegetale, per quanto riguarda il primo essendo la raccolta effettuata ogni 14 giorni e i recuperi di sabato, quindi se non vi è una corrispondenza di 14 giorni tra una data e l'altra allora non è un valore che deriva da un giro programmato. Le stesse operazioni sono state fatte per il rifiuto Vegetale, dove la raccolta ha una frequenza ogni 14 giorni per i mesi di Gennaio, Febbraio e Marzo mentre nei restanti mesi dell'anno è ogni 7 giorni ed inoltre sono presenti delle letture relative ai servizi cimiteriali.

Inoltre, per quanto riguarda il rifiuto Vegetale, si è voluto considerare un sottoinsieme di zone dove viene eseguita la raccolta. Nelle zone appartenenti a TV3 infatti negli anni la raccolta del rifiuto Vegetale veniva eseguita assieme al rifiuto Umido, e perciò non si terrà considerazione di queste zone.

Queste operazioni sono necessarie se le nostre analisi vogliono prendere in considerazione il parametro della densità di un contenitore per una determinato rifiuto in un determinato periodo dell'anno, infatti le letture cimiteriali ad esempio andrebbero a "sporcare" le letture relative ai giri di raccolta programmati e di conseguenza "sporcherrebbe" anche la % di esposizione dei contenitori in un determinato periodo dell'anno, non rendendo i dati effettivi.

Le operazioni di pulizia dei dati mirano a rendere la base dati concerne per quanto riguarda i giri di raccolta programmati per il rifiuto Vegetale e Secco non riciclabile.

4.7.2 Creazione dei dataset

Come visto ed effettuato nel dataset relativo alle rilevazioni nel periodo di 12 mesi, si importano i dati dall'ambiente Excel in R tramite la funzione `read_excel()` presente nel pacchetto `readxl`.

Successivamente vengono creati i due dataset relativi ai records di rifiuto Secco non riciclabile e rifiuto Vegetale, utilizzando la stessa sintassi utilizzata precedentemente per la creazione degli analoghi dataset nei 12 mesi.

Essendo i dati organizzati per rilevazioni mensili e per ogni comune nel bacino ai quali siamo interessati, si necessita di un raggruppamento mensile dei dati a nostra disposizione.

Procediamo creando due dataset, rispettivamente per il rifiuto Vegetale e Secco non riciclabile, aventi come dati delle medie mensili di densità, la % esposizione e la volumetria, mentre come somme le letture effettuate e i kg raccolti.

Come mostrato in Listing 4.6, viene effettuato un accorpamento di ogni parametro al quale siamo interessati ed successivamente viene creato un data frame tramite la funzione `data.frame()` rispettivamente per i dati accorpati di rifiuto Secco non riciclabile e Vegetale.

Listing 4.6: Raggruppamento dataset per mese.

```
DensEAcc <-Vegetale %>%
  group_by(Mese) %>%
  summarise(Densita_Mensile = mean('Densita_Mensile' (mensile)))
DensSAcc <-Secco %>%
  group_by(Mese) %>%
  summarise(Densita_Mensile = mean('Densita_Mensile' (mensile)))
.
.
.
LetEAcc <-Vegetale %>%
  group_by(Mese) %>%
  summarise(Letture_Mensile = sum(Letture))
LetSAcc <-Secco %>%
  group_by(Mese) %>%
  summarise(Letture_Mensile = sum(Letture))

SeccoAcc <- data.frame(Mese=DensSAcc$Mese,
  Densita=DensSAcc$Densita_Mensile,
  Esposizione=EspSAcc$Esp_Mensile,
  Volume=VolSAcc$Vol_Mensile,
  Letture=LetSAcc$Letture_Mensile,
  Litri=LitriSAcc$Litri_Mensile,
  kg=KgSAcc$Kg_Mensile)

VegetaleAcc <- data.frame(Mese=DensEAcc$Mese,
  Densita=DensEAcc$Densita_Mensile,
  Esposizione=EspEAcc$Esp_Mensile,
  Volume=VolEAcc$Vol_Mensile,
  Letture=LetEAcc$Letture_Mensile,
```

```
Litri=LitriEAcc$Litri_Mensile,
kg=KgEAcc$Kg_Mensile)
```

Sono stati trasformati allo stesso modo, i vettori di dati relativi a densità mensile, percentuale di esposizione mensile, media di volumetria mensile e kg raccolti.

4.7.3 Analisi

Si vuole ottenere una visualizzazione grafica delle seguenti serie storiche, mettendo in relazione le varie serie, per analizzare i dati e comprenderne la struttura e la relazione tra le variabili.

Edi nostro interesse, ed utile da esplorare le relazioni che intercorrono tra le letture effettuate e rispettivamente la densità mensile, la media di volumetria e infine la quantità di rifiuto in kg raccolta.

Dai grafici che mettono in relazione letture e densità mensile rispettivamente per il rifiuto Vegetale, mostrato in Figura 4.8 e il rifiuto Secco non riciclabile, mostrato in Figura 4.9, possiamo dedurre molteplici osservazioni.

Analizzando il grafico in Figura 4.8 relativo al rifiuto Vegetale, si nota come la serie appartenente alle letture ogni anno assuma una forma a campana avente concavità verso il basso, quindi è tendenzialmente in aumento a partire dall' inizio di ogni anno fino a subire un calo procedendo verso i mesi più freddi dell'anno.

Inoltre si può notare subito una possibile stagionalità nelle letture.

Per quanto riguarda la densità media mensile dei contenitori, si può notare come tendi ad aumentare in relazione alla tendenza in salita della serie relativa alle letture e questo ci porta a pensare che con l'aumento delle temperature, dovuto all'avvicinarsi dei mesi caldi, il contenuto dei contenitori da vuotare aumenti altrettanto. Questo fenomeno viene ricondotto al fatto che le utenze si predispongano alle attività di giardinaggio.

Analizzando il grafico in Figura 4.9 appartenente al rifiuto Secco non riciclabile si possono notare numeri picchi sia per la serie delle letture che per la serie della densità.

Per quanto riguarda la serie delle letture possiamo evidenziare quattro picchi in aumento, uno per ogni fine anno rispettivamente dicembre 2018, dicembre 2019, dicembre 2020, dicembre 2021. L'aumento delle letture nel mese di dicembre è un fenomeno riconducibile alla caratteristica del rifiuto Secco: lo svuotamento a pagamento. Quest'ultimo prevede che ogni utenza, per contratto, paghi preventivamente un numero di vuotamenti pari al numero di componenti nel nucleo familiare più uno ogni anno.

Si nota come ogni anno siano presenti altri picchi con la stessa frequenza tra i picchi dei mesi di dicembre, la maggior parte delle utenze quindi espone il contenitore mediamente 3/4 volte l'anno e, ipotizzando un nucleo familiare medio di 3/4 membri, si ottengono più vuotamenti di quanti mediamente si necessitano. Questo porta ad esporre, a fine anno, il contenitore anche se non contiene un quantitativo di rifiuto sostanzioso.

Notiamo, infatti come la serie relativa alla densità ad ogni mese di dicembre subisca un calo notevole.

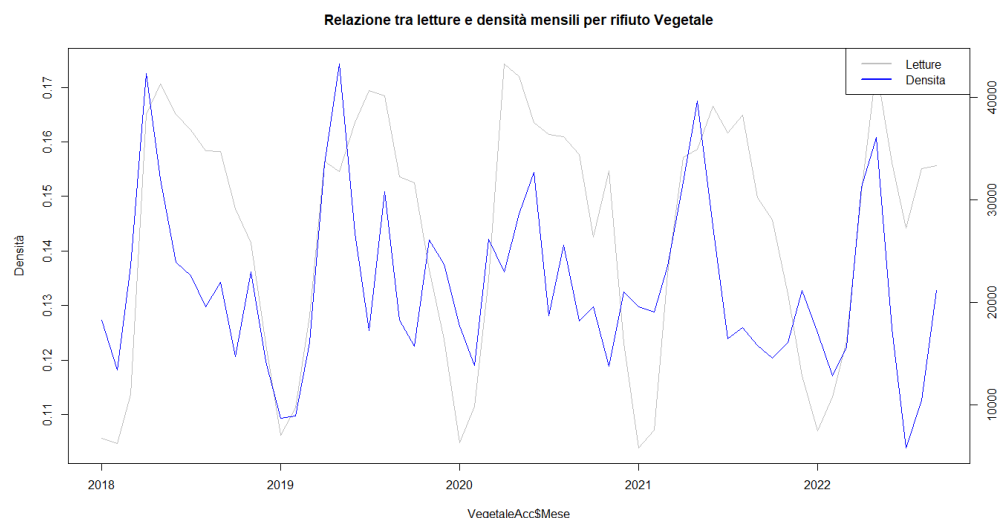


Figura 4.8: Grafico che mette in relazione due serie storiche (Letture e densità mensile) per il rifiuto Vegetale.

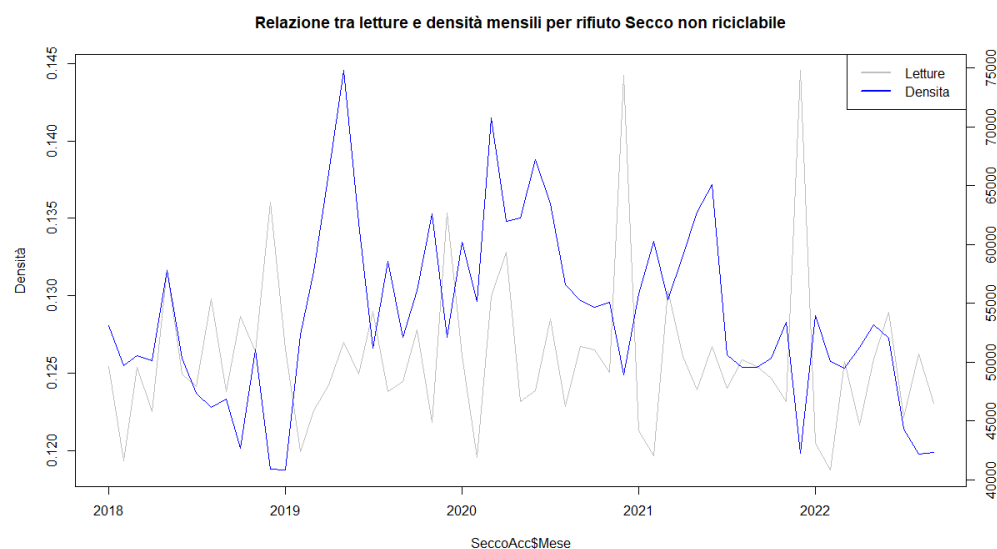


Figura 4.9: Grafico che mette in relazione due serie storiche (Letture e densità mensile) per il rifiuto Secco non riciclabile.

Analizzando il grafico appartenente al rifiuto Vegetale in Figura 4.10, si nota come la tendenza della serie relativa alla media di volumetria raccolta abbia una tendenza in calo, questo può significare che sempre più utenze preferiscono un contenitore più piccolo rispetto al contenitore da 240 Litri.

La stagionalità che si può osservare della serie storica relativa alla media di volumetria è caratterizzata da un aumento nei mesi finali di ogni anno e un calo andando verso i mesi primaverili ed estivi.

Questo è riconducibile al fatto che nei mesi freddi le utenze private non espongano quasi

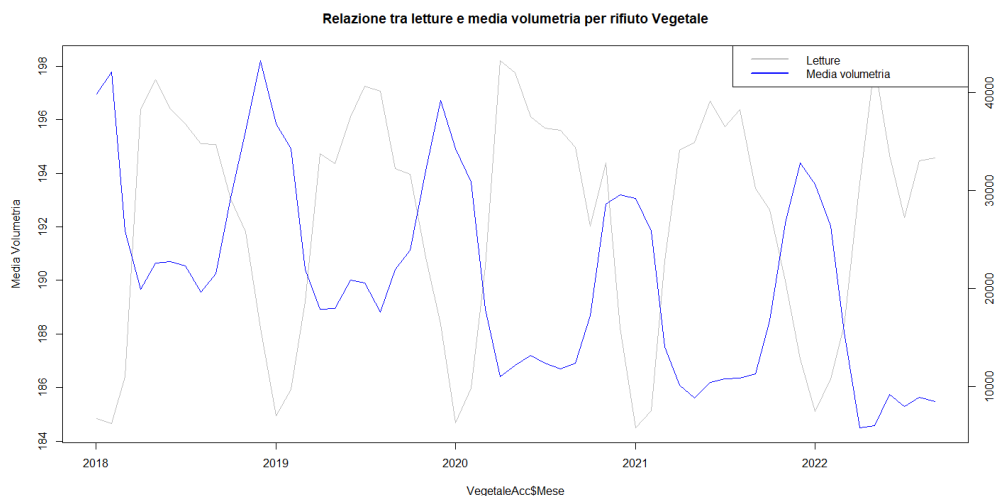


Figura 4.10: Grafico che mette in relazione due serie storiche (Letture e media volumetria mensile) per il rifiuto Vegetale.

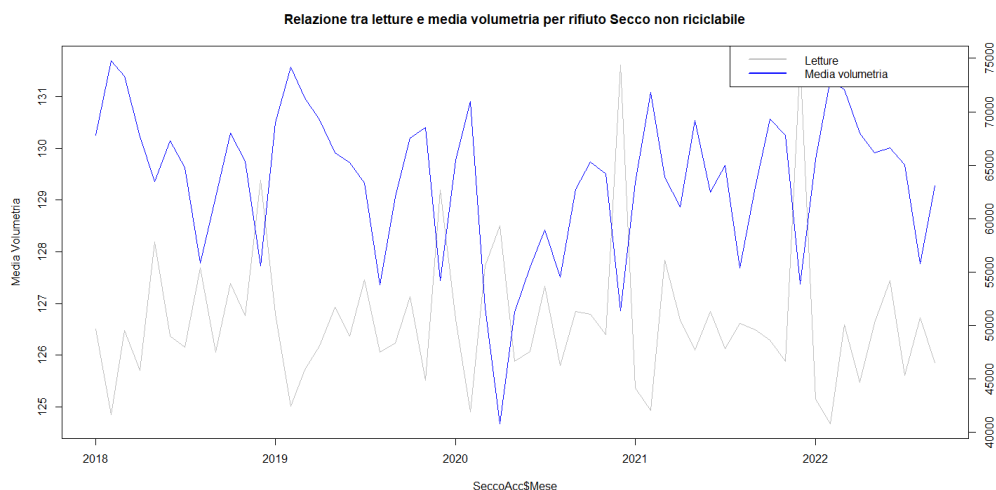


Figura 4.11: Grafico che mette in relazione due serie storiche (Letture e media volumetria mensile) per il rifiuto Secco non riciclabile.

mai il contenitore e quindi la volumetria media fa affidamento alle utenze pubbliche come i ristoranti, aziende che operano con determinati materiali, ecc. Quest'ultime, solitamente, hanno in dotazione un contenitore di alto volume.

Nei mesi più caldi, dove l'attività di giardinaggio possono esser svolte dalle utenze private, la media di volumetria è minore.

Per quanto riguarda il grafico del rifiuto Secco mostrato in Figura 4.11 non riciclabile si notano come i quattro picchi riconducibili ai mesi di dicembre di ogni anno, portino la media di volumetria ad abbassarsi.

In generale all'aumentare delle letture, si incorre in una diminuzione di densità di

rifiuto nel contenitore e questo avviene soprattutto nei mesi di dicembre in ogni anno.

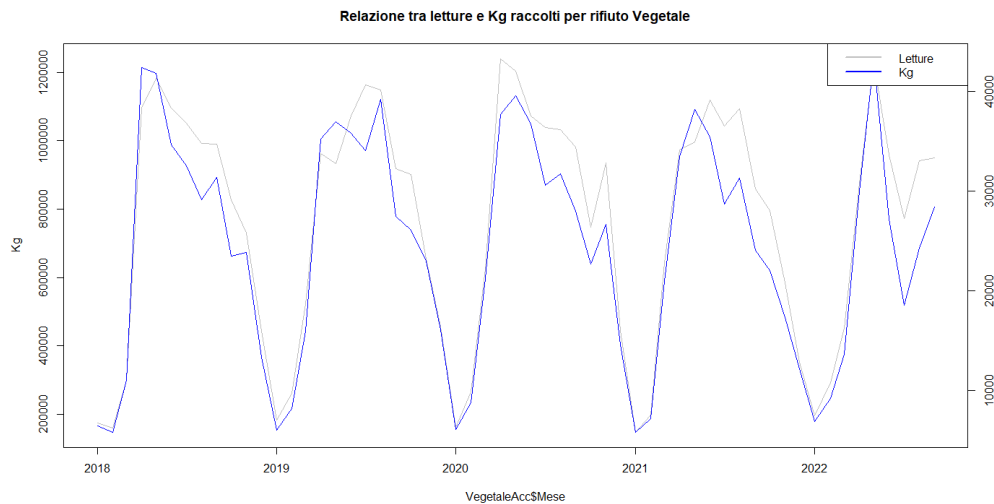


Figura 4.12: Grafico che mette in relazione due serie storiche (Lettture e Kg) per il rifiuto Vegetale.

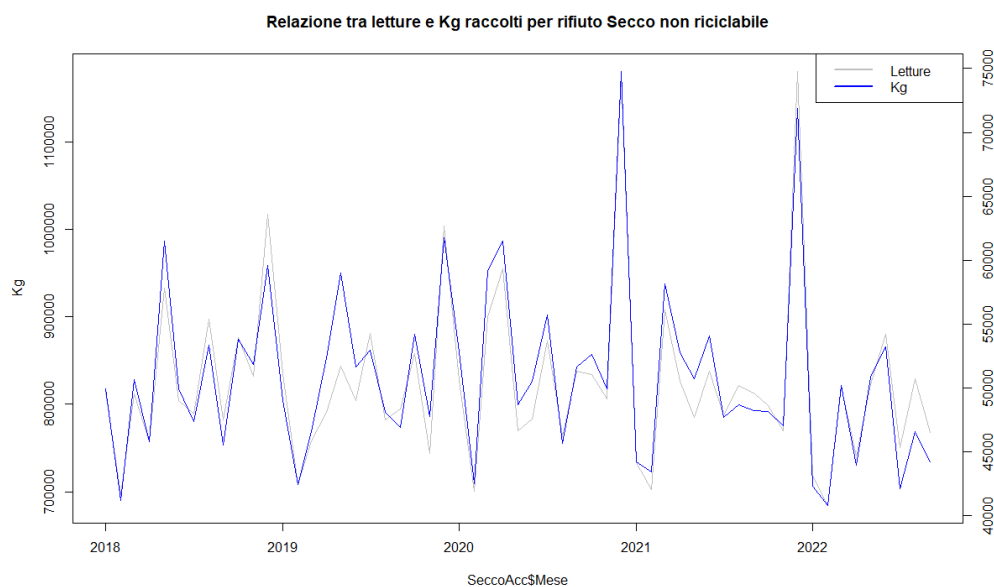


Figura 4.13: Grafico che mette in relazione due serie storiche (Lettture e Kg) per il rifiuto Secco non riciclabile.

Per quanto riguarda entrambi i grafici mostrati in Figura 4.12 e 4.13 si può notare come le letture seguano lo stesso andamento dei kg raccolti di entrambe le tipologie di rifiuto.

Le serie relative ai kg denotano quindi una potenziale stagionalità, come le rispettive

letture.

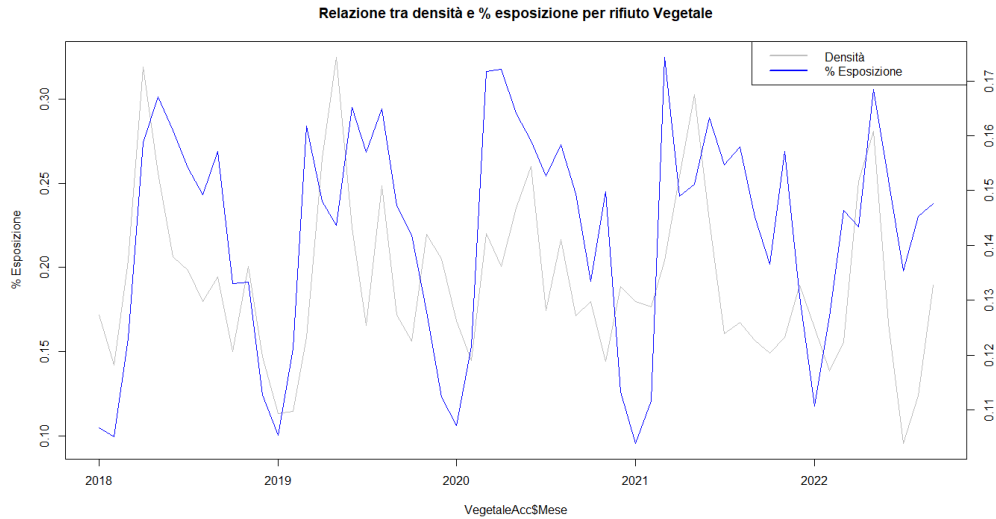


Figura 4.14: Grafico che mette in relazione due serie storiche (densità ed % esposizione) per il rifiuto Vegetale.

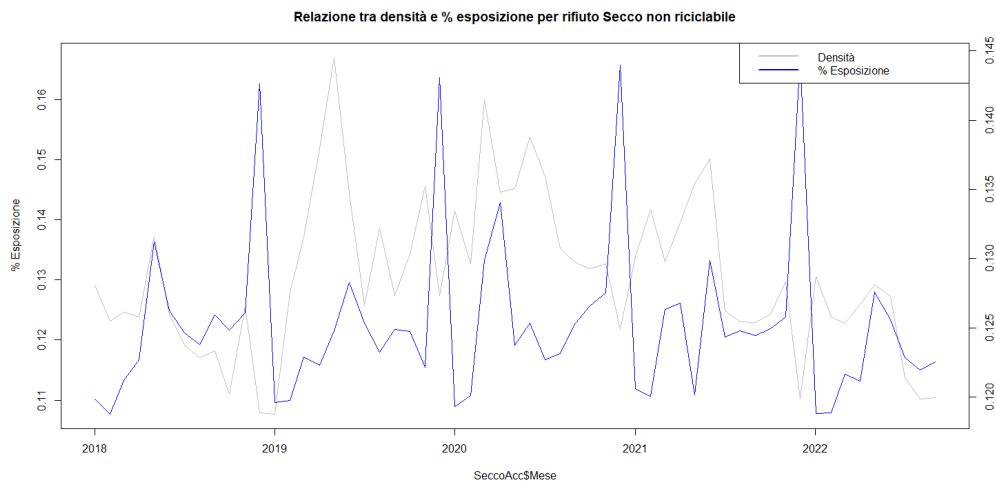


Figura 4.15: Grafico che mette in relazione due serie storiche (densità e % esposizione) per il rifiuto Secco non riciclabile.

Analizzando entrambi i grafici mostrati in Figura 4.14 e 4.15 notiamo che la % di esposizione di entrambi i rifiuti segue una stagionalità, dovuta al fatto che la serie abbia dei valori calcolabili tramite il numero di letture.

Di conseguenza la densità si comporterà come in precedenza descritto per quanto riguarda la relazione tra letture e densità.

Decomposizione della serie storica

La trasformazione di dati in un oggetto di serie storica tramite il comando *ts()* [13] in R è un passaggio importante per l'analisi delle serie temporali. Il comando trasforma un vettore di dati in un oggetto di serie temporale, che consente di specificare la frequenza dei dati e di utilizzare le funzioni e gli strumenti specifici per le serie temporali in R.

L'oggetto di serie temporale è un contenitore che mantiene informazioni sul tempo associato ai dati. Queste informazioni sono utilizzate da molte funzioni di R per elaborare e analizzare le serie temporali, come *stl()* e *decompose()* che utilizzano le informazioni sul tempo per effettuare la decomposizione della serie temporale in componenti come trend, stagionalità e rumore.

L'utilizzo di un oggetto di serie temporale è importante perché permette di utilizzare funzioni e strumenti specifici per le serie temporali, che altrimenti non sarebbero disponibili per i dati semplici in formato vettore.

Listing 4.7: Trasformazione di dati relativi alle letture in un oggetto di serie storica tramite il comando *ts()*

```
ts_letE<-ts(VegetaleAcc$Letture, start=c(2018,01,01), end=c
(2022,09,31), frequency = 12)
ts_letS<-ts(SeccoAcc$Letture, start=c(2018,01,01), end=c
(2022,09,31), frequency = 12)
```

Sono state eseguite le trasformazioni in oggetti di serie storica anche i dati relativi a densità, % di esposizione, media di volumetria e kg raccolti per entrambe le tipologie di rifiuto.

Al fine di analizzare le componenti stagionali, tendenziali e causali delle serie storiche appena trasformate, è possibile utilizzare l'analisi che prevede la separazione della serie storica in queste tre componenti distinte [3]:

- * Componente stagionale: rappresenta la componente periodica della serie storica, che si ripete regolarmente in un periodo di tempo specifico;
- * Componente tendenziale: rappresenta la componente di lungo termine della serie storica, che riflette la tendenza generale delle tendenze a lungo termine;
- * Componente causale: rappresenta la componente della serie storica che è causata da eventi esterni come cambiamenti nelle condizioni economiche o politiche.

Per analizzare queste componenti in R, è possibile utilizzare la funzione *stl()* (Seasonal Decomposition of Time Series) [13] del pacchetto "stats".

Si vogliono considerare in particolare determinati parametri. Sapendo che il campo densità e il campo %esposizione sono entrambi dei campi calcolati, cioè si possono ricavare dai dati presenti nella base dati, rispettivamente Kg/Litri e Letture/Contenitori Potenziali, ci si vuole concentrare sull'analisi di singoli parametri quali il numero di letture, la produzione in kg di rifiuto e la volumetria media.

Dai due grafici, mostrati in Figura 4.16 e 4.17 notiamo come il trend di entrambe le serie sia in tendenza di calo negli ultimi anni.

La componente stagionale nella serie relativa alle letture del rifiuto Vegetale si può notare come sia molto simile ai dati rilevati (grafico prima riga) e come la componente

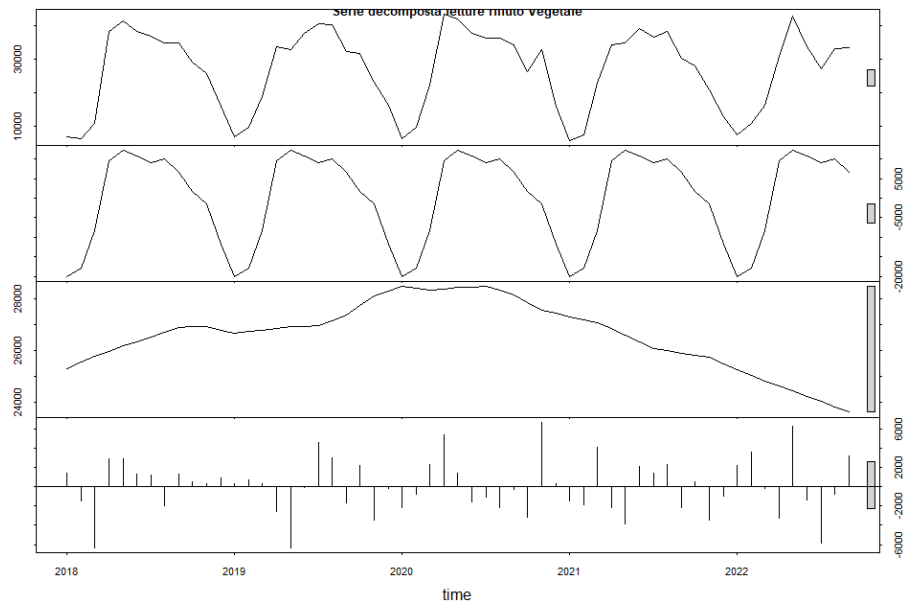


Figura 4.16: Decomposizione della serie storica relativa alle letture del rifiuto Vegetale.

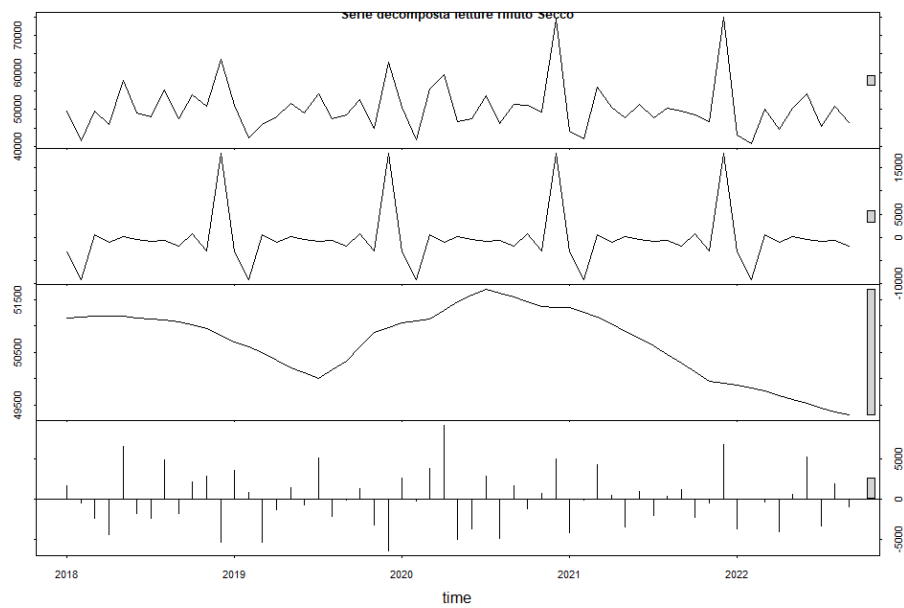


Figura 4.17: Decomposizione della serie storica relativa alle letture del rifiuto Secco non riciclabile.

di disturbo sia presente ma non tende ad influenzare troppo la componente stagionale. Invece la componente residua presente nelle letture del rifiuto Secco è più determinante

in base alle differenze dai primi due grafici.

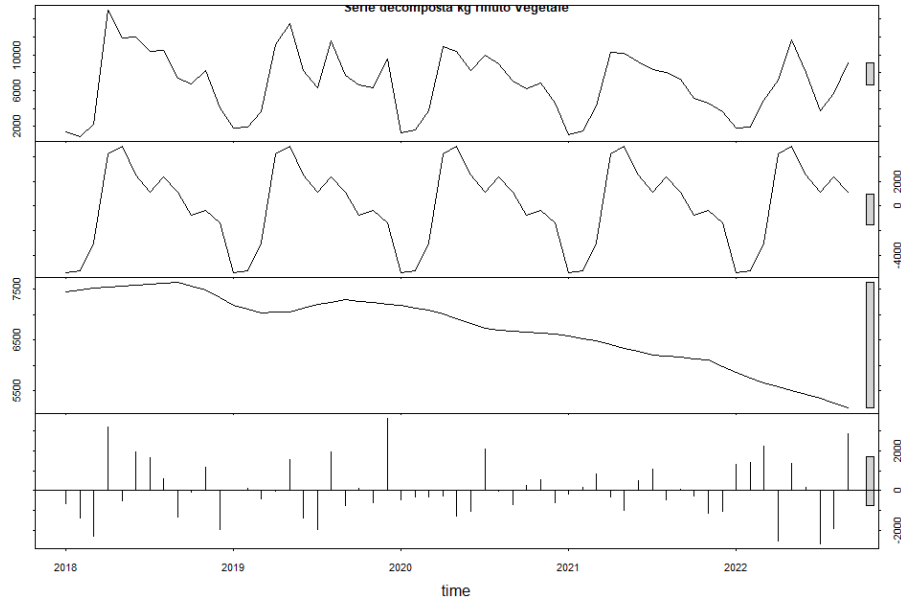


Figura 4.18: Decomposizione della serie storica relativa ai kg del rifiuto Vegetale.

Notiamo dai grafici in Figura 4.18 e 4.19 che la tendenza della produzione di rifiuto Vegetale in kg sia in calo, mentre la produzione del rifiuto Secco subisce negli anni un'oscillazione fino a subire una tendenza in aumento finale.

Nei grafici mostrati in Figura 4.20 e 4.21, nella decomposizione della serie appartenente sia al rifiuto Vegetale sia al rifiuto Secco, si evince un disturbo sostenuto nella stagionalità, con un trend in calo in entrambe le tipologie di rifiuto.

Si vuole inoltre poter determinare se le serie corrispondenti siano stazionarie oppure no. Per fare ciò, verifichiamo la stazionarietà dei dati residui tramite i test di Augmented Dickey-Fuller (ADF) e KPSS (Kwiatkowski-Phillips-Schmidt-Shin) [34] come mostrato in Listing 4.8.

Listing 4.8: Esecuzione test ADF e KPSS.

```
adf.test(stl_decomposed_letE$time.series[,3])

Augmented Dickey-Fuller Test

data: stl_decomposed_letE$time.series[, 3]
Dickey-Fuller = -5.0599, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

> kpss.test(stl_decomposed_letE$time.series[,3])
```

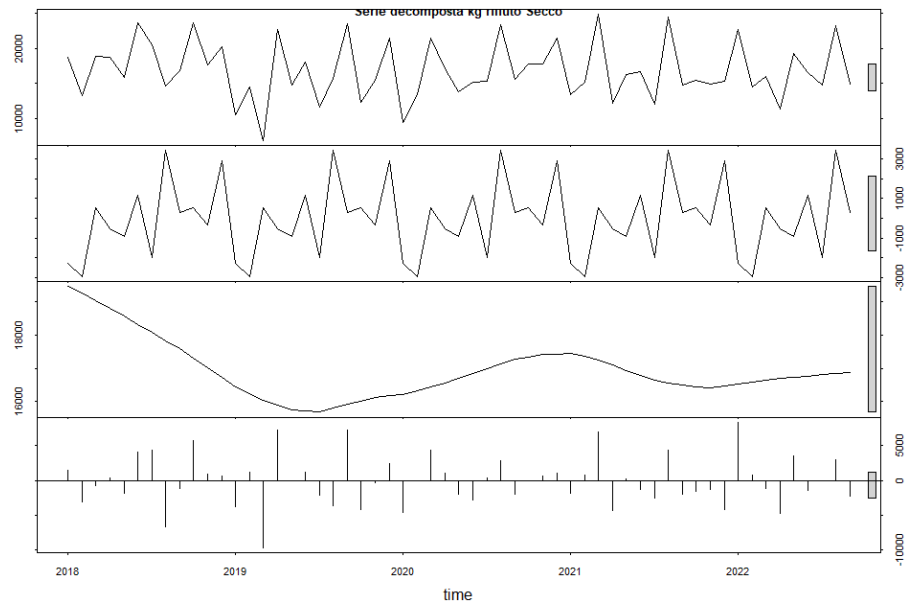


Figura 4.19: Decomposizione della serie storica relativa ai kg del rifiuto Secco non riciclabile.

```

      KPSS Test for Level Stationarity

data: stl_decomposed_letE$time.series[, 3]
KPSS Level = 0.028391, Truncation lag parameter = 3, p-value =
  0.1

> adf.test(stl_decomposed_letS$time.series[,3])

      Augmented Dickey-Fuller Test

data: stl_decomposed_letS$time.series[, 3]
Dickey-Fuller = -6.3387, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

> kpss.test(stl_decomposed_letS$time.series[,3])

      KPSS Test for Level Stationarity

data: stl_decomposed_letS$time.series[, 3]
KPSS Level = 0.024099, Truncation lag parameter = 3, p-value =
  0.1

> adf.test(stl_decomposed_kgE$time.series[,3])

      Augmented Dickey-Fuller Test

data: stl_decomposed_kgE$time.series[, 3]

```

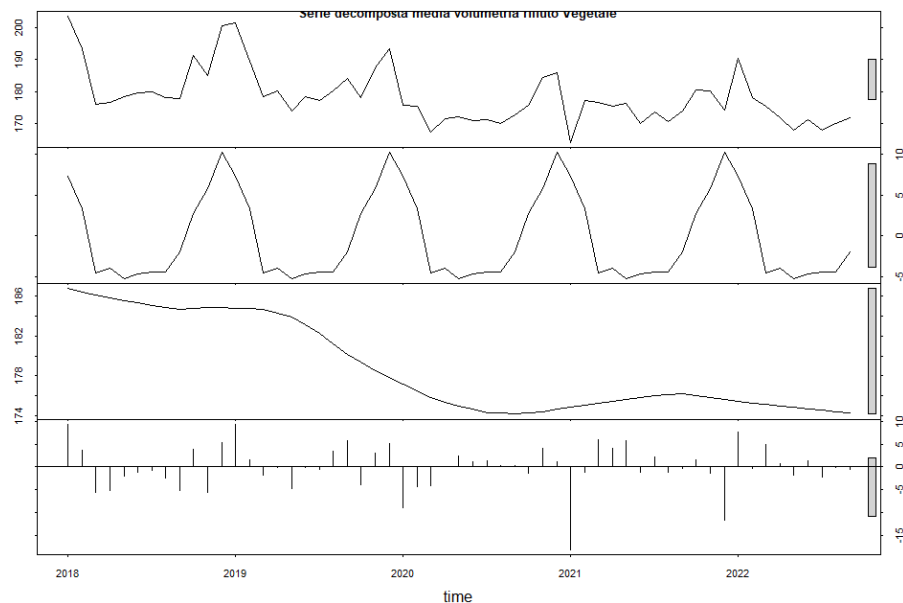


Figura 4.20: Decomposizione della serie storica relativa alla media di volumetria del rifiuto Vegetale.

```
Dickey-Fuller = -3.7948, Lag order = 3, p-value = 0.02486
alternative hypothesis: stationary

> kpss.test(stl_decomposed_kgE$time.series[,3])

      KPSS Test for Level Stationarity

data:  stl_decomposed_kgE$time.series[, 3]
KPSS Level = 0.029757, Truncation lag parameter = 3, p-value =
0.1
```

I risultati ottenuti, mostrati in Listing 4.8, indicano che i test di ADF sulle componenti stagionale e residuale confermano la stazionarietà.

Inoltre, i *p-values* ottenuto è inferiore a un livello di significatività del 1%. Questo suggerisce che l'ipotesi nulla che la serie è non-stazionaria può essere rifiutata. Questo significa che la serie è stazionaria o, in altre parole, che le sue proprietà statistiche non cambiano nel tempo.

I risultati ottenuti dai test KPSS, suggeriscono che l'ipotesi nulla che la serie è stazionaria non può essere rifiutata. Questo significa che la serie è stazionaria, ovvero che le sue proprietà statistiche non cambiano nel tempo.

Eseguendo un test della stazionarietà sulla componente residua di una serie storica, si può dedurre se la componente residua sia stazionaria o meno. Se la componente residua è stazionaria, significa che non ha una tendenza nel tempo

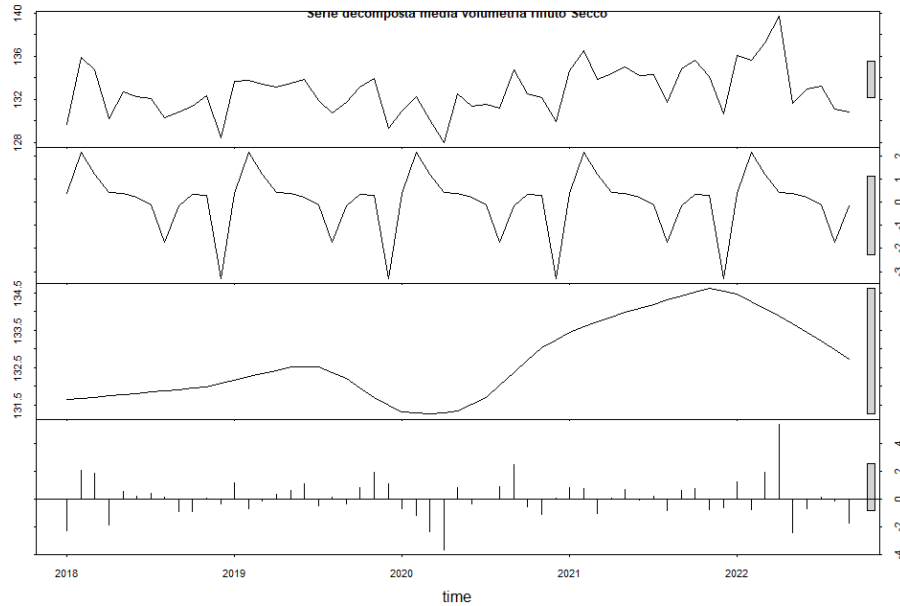


Figura 4.21: Decomposizione della serie storica relativa alla media di volumetria del rifiuto Secco non riciclabile.

e che le sue proprietà statistiche non cambiano. Questo è importante perché una componente residua non stazionaria potrebbe influire negativamente sulle previsioni.

4.8 Modello predittivo

Lo sviluppo di un modello per previsioni future è un processo che utilizza dati storici per prevedere eventi futuri. Questo è un compito importante in molti settori, come ad esempio finanza, marketing, logistica e previsioni meteorologiche. Il modello è basato su un'analisi dei dati storici e sulle relazioni tra le variabili che influenzano gli eventi futuri. Una volta che il modello è stato costruito e verificato con successo sui dati storici, può essere utilizzato per fare previsioni future, fornendo così una base informata per prendere decisioni a livello di gestione [2], [18].

La scelta del modello giusto dipende dalla natura dei dati e dal tipo di previsione che si vuole effettuare. Ad esempio, un modello di regressione lineare può essere adatto per prevedere una relazione lineare tra due variabili. Una volta che il modello è stato sviluppato, è importante verificarne la validità e la precisione utilizzando dati storici e confrontandoli con i dati reali [23].

In definitiva, lo sviluppo di un modello per previsioni future può fornire un grande valore per le aziende e le organizzazioni, poiché può aiutare a prevedere il futuro e a prendere decisioni informate in modo più efficiente.

Si vuole valutare come l'esposizione possa aumentare o diminuire in base all'andamento dei kg, all'andamento dei litri, e all'andamento della media di volumetria. Si vuole valutare graficamente, come mostrato in Figura 4.22 per quanto riguarda il rifiuto Secco, la distribuzione di entrambe le tipologie di rifiuto per le letture perchè l'esposizione è direttamente proporzionale alle letture, ed il risultato visualizzato si avvicina ad una distribuzione normale.

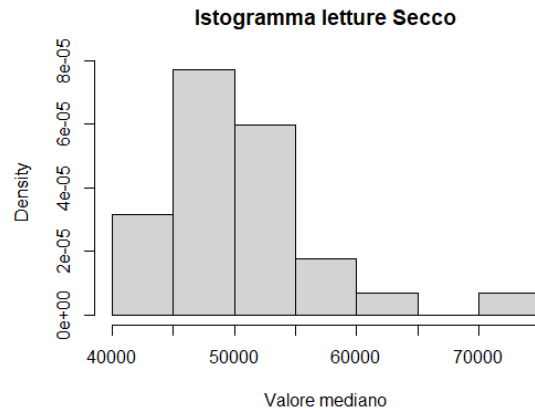


Figura 4.22: Istogramma distribuzione letture rifiuto Secco non riciclabile.

La regressione lineare è un modello matematico che descrive la relazione tra una o più variabili indipendenti (chiamate anche "spiegazioni") e una variabile dipendente (o risultato) che si desidera prevedere. Il modello è chiamato "lineare" perché descrive la relazione tra le variabili come una linea retta o, più in generale, un iperpiano.

L'obiettivo della regressione lineare è quello di trovare la linea che descrive la relazione più stretta possibile tra la variabile dipendente e le variabili indipendenti. La linea di regressione viene calcolata utilizzando i coefficienti del modello, che descrivono la pendenza e l'intercetta della linea [1].

La regressione lineare è una tecnica statistica consolidata e le aziende la utilizzano per convertire in modo affidabile e prevedibile i dati non elaborati in business intelligence e informazioni fruibili [11].

Si vuole provare a stimare un modello di regressione lineare semplice della forma:

$$topredict = \beta_0 + \beta_1 variable + \epsilon$$

Dove *topredict* e *variable*, identificano rispettivamente la variabile da predire e la variabile (o le variabili) utilizzate per la previsione.

La funzione *lm()* in R è una delle funzioni più utilizzate per creare un modello di regressione lineare, questa funzione prevede che i dati siano organizzati come un data frame, con le variabili indipendenti sulla sinistra e la variabile dipendente sulla destra. La sintassi generale per eseguire la regressione lineare con la funzione *lm()* è mostrata in Listing 4.9.

Listing 4.9: Creazione modello con `lm()`.

```
modelloE <- lm(Letture ~ kg + Litri + Volume, data=VegetaleAcc)
modelloS <- lm(Letture ~ kg + Litri + Volume, data=SeccoAcc)
```

- * "Letture" rappresenta la variabile dipendente che si desidera prevedere.
- * "kg + Litri + Volume" rappresentano le variabili indipendenti che influenzano la variabile dipendente.
- * "data = dati" specifica il data frame che contiene i dati che verranno utilizzati per la regressione lineare.
- * "modello" è il nome che verrà assegnato al modello di regressione lineare.

Si vuole visualizzare i risultati della regressione lineare. Sfruttiamo la funzione `summary()` la quale fornisce una panoramica completa del modello di regressione lineare che è stato creato, comprese le informazioni sulla qualità del modello [1].

L'output, mostrato in Listing 4.10, fornisce un oggetto (modello) contenente moltissime informazioni.

Listing 4.10: Informazioni modelli tramite `summary()`.

```
> summary(modelloE)

Call:
lm(formula = Letture ~ kg + Litri + Volume, data = VegetaleAcc)

Residuals:
    Min       1Q   Median       3Q      Max
-327.93 -106.16  -19.62   80.57  424.14

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.345e+04  1.642e+03  14.282  <2e-16 ***
kg           2.939e-04  2.657e-04   1.106    0.274
Litri        5.220e-03  4.146e-05 125.885  <2e-16 ***
Volume      -1.220e+02  8.366e+00 -14.581  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
                 0.1 ' ' 1

Residual standard error: 164.9 on 53 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9998
F-statistic: 9.601e+04 on 3 and 53 DF,  p-value: < 2.2e-16

> summary(modelloS)

Call:
lm(formula = Letture ~ kg + Litri + Volume, data = SeccoAcc)

Residuals:
    Min       1Q   Median       3Q      Max
-221.54  -85.64  -25.10   86.12  285.33
```

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.215e+04  1.763e+03  29.576  <2e-16 ***
kg           -2.130e-04  5.085e-04  -0.419    0.677
Litri        7.783e-03  6.070e-05  128.225  <2e-16 ***
Volume      -4.060e+02  1.295e+01  -31.356  <2e-16 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'
                 0.1  ' '  1

Residual standard error: 116.4 on 53 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 6.145e+04 on 3 and 53 DF,  p-value: < 2.2e-16

```

La funzione `summary()` fornirà informazioni come [1]:

- * Coefficienti di regressione: questi sono i valori che descrivono l'impatto di ogni variabile indipendente sul valore della variabile dipendente;
- * P-valori: questi sono i valori che descrivono la probabilità che i coefficienti di regressione siano zero, ovvero che la variabile indipendente non abbia alcun effetto sulla variabile dipendente;
- * R^2 (coefficiente di determinazione): questo valore descrive la percentuale di variazione nella variabile dipendente che può essere spiegata dal modello di regressione lineare. Un valore di R^2 vicino a 1 indica una buona qualità del modello, mentre un valore vicino a 0 indica una qualità del modello scarsa;
- * $Adj.R^2$: questo valore è simile al coefficiente di determinazione R^2 , ma tiene conto del numero di variabili indipendenti nel modello;
- * Residui: questi sono i valori che descrivono la differenza tra il valore effettivo della variabile dipendente e il valore previsto dal modello di regressione lineare.

Possiamo notare come entrambi i coefficienti di determinazione siano molto vicini ad 1, e lo stesso vale per il coefficiente $Adj.R^2$ quindi possiamo definire entrambi i modelli aventi una qualità molto buona.

Si nota inoltre la presenza di asterischi a fianco dei parametri. Questi asterischi determinano l'importanza delle variabili, e notiamo in particolare come i kg non siano così rilevanti non avendo asterischi corrispondenti.

Si vuole dunque riformulare il modello senza i kg, come mostrato in Listing 4.11.

Listing 4.11: Creazione modello con `lm()`.

```

modelloE <- lm(Letture ~ Litri + Volume, data=VegetaleAcc)
modelloS <- lm(Letture ~ Litri + Volume, data=SeccoAcc)

```

e si ottiene, riproducendo la funzione `summary()` sui nuovi modelli, l'output mostrato in Listing 4.12.

Listing 4.12: Informazioni modelli tramite `summary()`.

```

> summary(modelloE)

```

```

Call:
lm(formula = Letture ~ Litri + Volume, data = VegetaleAcc)

Residuals:
    Min       1Q   Median       3Q      Max
-321.42 -115.90   -3.51    98.84   464.06

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.336e+04  1.643e+03   14.21  <2e-16 ***
Litri        5.263e-03  1.411e-05   372.93  <2e-16 ***
Volume     -1.216e+02  8.374e+00   -14.52  <2e-16 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'
                 0.1      1

Residual standard error: 165.2 on 54 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9998
F-statistic: 1.434e+05 on 2 and 54 DF,  p-value: < 2.2e-16

> summary(modelloS)

Call:
lm(formula = Letture ~ Litri + Volume, data = SeccoAcc)

Residuals:
    Min       1Q   Median       3Q      Max
-226.35  -87.34  -26.68   83.57  282.45

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.201e+04  1.715e+03   30.33  <2e-16 ***
Litri        7.759e-03  2.201e-05   352.50  <2e-16 ***
Volume     -4.050e+02  1.264e+01   -32.04  <2e-16 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'
                 0.1      1

Residual standard error: 115.5 on 54 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 9.361e+04 on 2 and 54 DF,  p-value: < 2.2e-16

```

Si vuole ora, prevedere dei valori per entrambi i modelli.

La funzione `predict()` [13] in R viene utilizzata per prevedere i valori futuri della variabile dipendente. La sintassi generale per utilizzare la funzione `predict()` è mostrata in Listing 4.13:

Listing 4.13: Sintassi `predict()`.

```

predictionsE <- predict(modelloE, newdataE)
predictionsS <- predict(modelloS, newdataS)

```

Le variabili `newdataE`, `newdataS` rappresentano il data frame che contiene le variabili indipendenti per le quali si desidera prevedere il valore della variabile dipendente. Nel nostro caso, gli oggetti `newdataE` e `newdataS` contengono valori di kg da Novembre

2021 fino ad Ottobre 2022, visualizzati con rilevazioni mensili.

Il risultato della funzione sarà un vettore di valori previsti, dove ogni valore rappresenta la previsione del valore della variabile dipendente per le corrispondenti righe del data frame *newdata* corrispondente.

È importante notare che il data frame *newdata* deve avere le stesse colonne che sono state utilizzate nel modello di regressione lineare, altrimenti la funzione *predict()* non funzionerà. Inoltre, è importante che i valori delle variabili indipendenti nel data frame "newdata" siano validi e sensati, in modo che le previsioni siano affidabili.

Si vuole, ora, mettere in relazione graficamente i valori stimati sulla base del modello e la retta stimata ai minimi quadrati. I comandi per visualizzare il grafico è mostrato in Listing 4.14 per quanto riguarda il rifiuto Secco, analogamente sono stati eseguiti gli stessi comandi per il rifiuto Vegetale.

Listing 4.14: Comandi per relazione valori stimati e retta minimi quadrati

```
plot(SeccoAcc$kg, SeccoAcc$Letture, pch=19, cex=0.5,
      xlab='Kg',
      ylab='Letture mediano')
## aggiungiamo i valori stimati
points(SeccoAcc$kg, valoriS.stimati, pch='x', col='green')
## aggiungiamo la retta stimata ai minimi quadrati
abline(coef(modelloS)[1], coef(modelloS)[2], lty=2, col='red',
        lwd=3)
```

Dalle figure 4.23 e 4.24, si può notare come la retta dei minimi quadrati e i punti stimati in verde seguono lo stesso andamento.

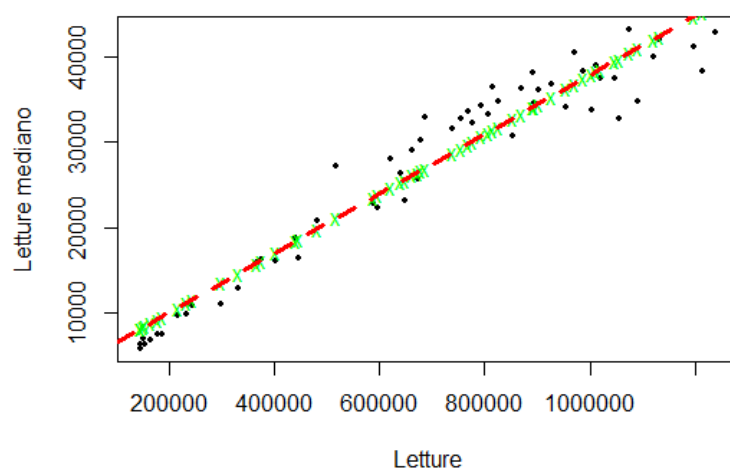


Figura 4.23: Dati osservati, valori stimati sulla base del modello e retta stimata ai minimi quadrati rifiuto Vegetale

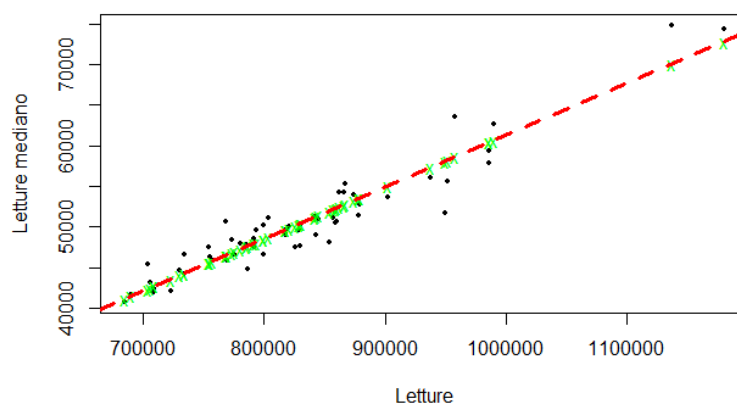


Figura 4.24: Dati osservati, valori stimati sulla base del modello e retta stimata ai minimi quadrati rifiuto Secco

Capitolo 5

Riorganizzazione Operativa del servizio PAP

La riorganizzazione del piano operativo è un processo continuo che mira a migliorare l'efficienza e l'efficacia delle attività aziendali.

In questo senso, piccoli accorgimenti possono fare la differenza e portare a notevoli miglioramenti.

L'adozione di nuove tecnologie e l'utilizzo di metodologie più efficienti possono anche contribuire a rendere il processo più snello e a supportare i dipendenti nello svolgimento del loro lavoro.

In definitiva, la riorganizzazione del piano operativo è un passo fondamentale per garantire il successo a lungo termine dell'azienda e per mantenere un vantaggio competitivo.

In particolare si vuole riorganizzare il piano operativo generato dall'algoritmo in grado di valutare ed indicare quante risorse sono necessarie, in base a vari parametri, per effettuare e completare una giornata di raccolta programmata.

5.1 Algoritmo attuale

L'algoritmo, in utilizzo fino ad oggi, tiene conto di determinati parametri fissati nel tempo e non dinamici.

Si fa l'esempio dell'esposizione, infatti è impossibile avere una % di esposizione costante per tutto l'anno, così vale anche per la densità del contenitore stesso.

In Figura 5.1 viene mostrata una parte dell'interfaccia, in Excel, dell'algoritmo utilizzato.

L'algoritmo calcola, sulla base dei parametri, le risorse necessarie per effettuare la raccolta per il rifiuto Vegetale. L'azienda utilizza un foglio Excel per effettuare questi calcoli, più in particolare, come si nota da Figura 5.1, esso è composto da una tabella avente come colonne i dati descritti in Sezione 3.1

Ma in forza di quanto detto in Sezione 2.6.1, per il rifiuto Vegetale si ha una frequenza di raccolta di $1v$ /settimana nel periodo da aprile a dicembre, mentre la frequenza diminuisce nei mesi di gennaio, febbraio e marzo rispettivamente di $1v/2$ settimane. Dal grafico in Figura 4.9, si evidenziano come nel mese di dicembre e marzo vi è una

ALGORITMO DI CALCOLO
DEFINIZIONE ORGANICO SQUADRE SERVIZIO PORTA A PORTA
VERSIONE 2.1

ZONA	ANTALM.	FATTORE F.	COMUNE	RIFIUTO	VOLUMETRIA (M3)	FREQUENZA (GIU' ANNO)	UTENZE TOTALI	S. ESPOSIZIONE	UTENZE EFFETTIVE	ONE TEMPI	TEMPO TRASFERIMENTO	NUMERO SCARICHI MESE	TEMPO SINGOLO SCARICO	PAUSA	OPERAZIONI ORGANICHE	TEMPO EFFETTIVO NETTO	NUMERO CONTENITORI SERVITI PER GIORNO	PRESE EFFETTIVE PER GIORNO	ESITO
ROSA	4.471	2	ARCADE	VEGETALE	240	1	304	35	306	7,60	0,05	3	0,40	0,50	0,25	5,10	30	63	0,7
GIALLA	7.841	3	BREDA DI PIAVE	VEGETALE	240	1	794	35	278	7,60	0,58	3	0,75	0,50	0,25	4,02	30	121	2,3
GIALLA	11.179	2	CARBONERA	VEGETALE	240	1	1372	35	480	7,60	0,58	3	0,75	0,50	0,25	4,02	30	121	4,0
ROSSA	12.332	2	CASALE SUL SILE	VEGETALE	240	1	2.030	35	711	7,60	1,33	3	0,50	0,50	0,25	4,02	30	121	5,3
ROSSA	11.304	2	CASIER	VEGETALE	240	1	1.053	35	546	7,60	0,31	3	0,25	0,50	0,25	5,19	23	91	3,6
ROSSA	5.102	3	GIAVERA DEL MONTELLO	VEGETALE	240	1	238	35	94	7,60	0,68	3	0,75	0,50	0,25	3,34	30	110	0,9
GIALLA	9.357	3	MASERADA SUL PIAVE	VEGETALE	240	1	1.023	35	380	7,60	0,33	3	0,75	0,50	0,25	4,27	30	120	2,8
BLU	4.250	3	MONASTEREN DI TREVISO	VEGETALE	240	1	500	35	175	7,60	0,30	3	0,58	0,50	0,25	4,21	30	126	1,4
VERDE	4.401	3	MORGANO	VEGETALE	240	1	399	35	140	7,60	0,95	3	0,48	0,50	0,25	4,46	30	134	1,0
ROSSA	6.841	4	NERVESA DELLA BATTAGLIA	VEGETALE	240	1	362	35	127	7,60	0,66	3	0,41	0,50	0,25	4,36	25	124	1,0
VERDE	21.996	2	PAESE	VEGETALE	240	1	2.940	35	1.032	7,60	0,70	3	0,41	0,50	0,25	4,32	30	140	7,0
CELESTE	12.495	2	PONCANO VENETO	VEGETALE	240	1	1.175	35	411	7,60	0,50	3	0,31	0,50	0,25	5,42	30	103	2,5
CELESTE	5.255	3	POVEGLIANO	VEGETALE	240	1	391	35	137	7,60	0,56	3	0,28	0,50	0,25	5,45	30	104	0,8
ROSSA	10.098	2	PREGANZOL	VEGETALE	240	1	2.549	35	892	7,60	1,00	3	0,25	0,50	0,25	5,10	30	103	5,8
VERDE	9.856	2	QUINDO DI TREVISO	VEGETALE	240	1	394	35	248	7,60	0,63	3	0,25	0,50	0,25	5,27	30	106	2,2
BLU	14.293	3	RONCADE	VEGETALE	240	1	1.740	35	609	7,60	0,85	3	0,41	0,50	0,25	4,77	30	143	4,3
BLU	13.032	3	SAN BIAGIO DI CALLALTA	VEGETALE	240	1	1.284	35	449	7,60	0,31	3	0,41	0,50	0,25	4,71	30	141	3,2
ROSSA	10.137	2	SILEA	VEGETALE	240	1	1.102	35	386	7,60	0,31	3	0,58	0,50	0,25	4,20	30	126	3,1
GIALLA	12.020	2	SPIRESANO	VEGETALE	240	1	1.297	35	454	7,60	0,33	3	0,75	0,50	0,25	4,27	30	128	3,5
ROSSA	11.951	3	SUSEGANA	VEGETALE	22	2	1.125	35	397	7,60	0,68	3	0,41	0,50	0,25	4,36	30	149	2,7
ROSSA	94.602	1	TREVISO - ZONA 3	VEGETALE	240	1	7.727	35	2.704	7,60	0,31	3	0,41	0,50	0,25	4,71	30	141	19,1
CELESTE	10.009	2	VILLORBA	VEGETALE	240	1	2.323	35	813	7,60	0,38	3	0,25	0,50	0,25	5,74	30	172	4,7

Figura 5.1: Base algoritmo operativo utilizzato da Contarina SpA.

corrispondenza rispettivamente di scarico e carico per quanto riguarda le letture e i litri.

Questo porta a ragionare sulla quantità, infatti si nota il picco massimo nei mesi da aprile a novembre e il picco minimo nei mesi tra gennaio e febbraio, evidenziando una tendenza in decrescita nel mese di dicembre e una tendenza in crescita nel mese di marzo.

Possiamo quindi concludere che la frequenza di raccolta andrebbe modificata, ponendo al mese di dicembre una frequenza di 1v/settimana essendo un mese in tendenza di scarico, e al mese di aprile una frequenza di 1v/2settimane essendo un mese in tendenza di carico.

Per il rifiuto Secco, la media di letture mensile tra il periodo di gennaio e novembre è minore rispetto alla media di letture del solo mese di dicembre, che vede un aumento di letture rispetto alla media dei restanti mesi dell'anno come mostrato in Figura 4.10.

Le utenze che posseggono il contenitore per il rifiuto Vegetale possono scegliere la volumetria più adatta alle loro esigenze (120 o 240 o 360). Valutando la media mensile del rapporto tra i litri e le letture del rifiuto Vegetale, si riesce ad ottenere una media mensile della dimensione dei contenitori vuotati in quel mese dell'anno.

Analizzando i grafici in Figura 4.9 e 4.10, si nota come nei periodi dove le letture hanno una tendenza in decrescita la volumetria dei contenitori aumenti, mentre decresce nei periodi interessati ad un aumento delle letture.

Il problema della frequenza di raccolta è uno degli aspetti principali per i quali si vuole aggiornare l'algoritmo, infatti si vogliono poter impiegare un numero di risorse corrette per le giornate di raccolta.

A tal scopo la riorganizzazione avviene su base mensile.

In pratica, l'algoritmo viene mantenuto, viene aumentata la sua granulosità. Infatti i calcoli avvengono esattamente come nell'algoritmo originale, cambia solamente la rilevazione.

Come si può notare da Figura 5.2, ogni riga rappresenta un determinato comune, sono

Figura 5.2: Parte nuovo algoritmo operativo utilizzato da Contarina SpA per il rifiuto Secco.

A tal scopo vengono calcolate le somme rispettivamente dei parametri da sommare per avere un dato mensile, come le letture, e le medie relative ad esempio alla durata dei giri di raccolta in un quel determinato mese e in quel determinato comune. Successivamente l'algoritmo calcolerà i parametri calcolabili come ad esempio il numero prese per ora.

Infine si ottiene un numero di risorse calcolato per ogni mese e per ogni comune. Nell'immagine viene visualizzato il risultato per quanto riguarda il rifiuto Secco, e si può notare come effettivamente il numero di risorse necessarie nel mese di Dicembre (un mese "critico" per questo rifiuto) siano maggiori rispetto a tutto l'anno. Ovviamente, nel caso si voglia mantenere il calcolo originale si avrà un sovraccarico di lavoro in base a quanto previsto dall'algoritmo originale "statico". Con i risultati ottenuti, invece, si cerca di ammortizzare il carico di lavoro, cercando un maggiore equilibrio.

Le operazioni sono state eseguite rispettivamente per il rifiuto Vegetale, caratterizzando il risultato come per il rifiuto Secco non riciclabile.

Capitolo 6

Conclusioni

6.1 Risultati ottenuti

La tesi ha avuto come scopo l'analisi di dati provenienti da rilevazioni di un anno (12 mesi) e provenienti da dati con rilevazioni mensili su 5 anni.

Per ottenere i risultati, abbiamo dapprima raccolto i dati e successivamente analizzati tramite il linguaggio R.

Tali analisi hanno prodotto una buona comprensione del servizio di raccolta porta a porta di Contarina SpA, riuscendo a comprendere a fondo determinati parametri che lo governano, con annesse relazioni tra di essi.

Successivamente sono stati analizzati i dati nel periodo temporale più lungo, andando ad evidenziare la presenza di eventuali trend e/o stagionalità nelle serie temporali.

Lo scopo nell'evidenziare la presenza di quest'ultimi voleva portare ad uno sviluppo di un modello per valutare la previsione delle letture tramite i valori dei kg, dei litri e della media di volumetria rispettivamente per il rifiuto Secco e Vegetale.

Lo scopo è stato raggiunto con risultati abbastanza soddisfacenti e da intendere come una buona base di partenza per impieghi futuri del modello.

Si ricorda, infatti, che lo sviluppo di tali modelli non preclude lo sviluppo di altrettanti modelli predittivi per altrettanti determinati parametri differenti da quelli esaminati nella tesi.

6.2 Raggiungimento degli obiettivi

Nella Tabella 1.1, in Sezione 1.3.2 vengono elencati gli obiettivi preliminari previsti dall'attività di *stage*, con i relativi risultati raggiunti.

Come si può notare da Tabella 6.1, gli obietti obbligatori sono stati soddisfatti, mentre non sono stati soddisfatti i vincoli desiderabili e facoltativo. Per quanto riguarda quest'ultimo, non c'è stato tempo per effettuare un approfondimento dei sistemi informatici in uso dall'azienda, avendo un limitato numero di ore rimanenti si è deciso di utilizzarle per sviluppare al meglio gli obiettivi obbligatori.

Il poco tempo a disposizione è stata conseguenza di problematiche sorte durante le fasi di estrazione e analisi dei dati.

Con certezza, quindi, non si sarebbero potuti iniziare e portare a termine gli obiettivi

Tabella 6.1: Tabella degli obiettivi consuntivati.

Descrizione obiettivo	Codice obiettivo	Risultato
Comprensione dell'architettura As Is	O01	Soddisfatto
Esecuzione analisi dati e caratterizzazione delle variabili	O02	Soddisfatto
Elaborazione di una proposta di riorganizzazione operativa del servizio PAP	O03	Soddisfatto
Esecuzione Value Stream Map	D01	Non Soddisfatto
Individuazione del set di dati da passare al data warehouse	D02	Non Soddisfatto
Valutazione dei sistemi informatici attualmente in uso e spunti di miglioramento	F01	Non Soddisfatto

non obbligatori.

6.3 Consuntivo delle tempistiche

La Tabella 6.2 riassume le attività preventivate, le relative ore a consuntivo e viene specificata la differenza di ore tra parentesi.

Dalla Tabella 6.2 si può notare come, rispetto alla Tabella 1.2 in Sezione 1.3.2, la gestione delle tempistiche e delle ore non sia allineata con i dati preventivati. In particolare, si sono risparmiate delle ore nell'attività di strategia di analisi grazie all'intenso approfondimento nelle attività di comprensione dello stato dell'arte. Tuttavia, nell'attività di raccolta e studio dati si vedono aumentare notevolmente le ore, questo è dovuto principalmente a problemi legati al reperimento e all'estrazione in azienda, che hanno allungato notevolmente le tempistiche. L'attività di analisi critica dei dati, invece, ha subito un aumento delle ore rispetto al preventivo dovuto al tempo richiesto per approfondire i concetti relativi allo studio delle serie storiche, modelli predittivi e il linguaggio R.

Tabella 6.2: Tabella delle attività pianificate e ore consuntivate.

Descrizione attività	Ore consuntivate
Inserimento in azienda	38
Comprensione dello stato dell'arte	38
Strategia di analisi	20 (-18)
Raccolta e studio dei dati	114 (+38)
Analisi critica dei dati	48 (+10)
Riorganizzazione	30 (-8)
Attività conclusive	20 (-18)
Totale ore	328 (+4)

6.4 Conclusioni

Lo scopo dello *stage* e gli obiettivi obbligatori sono stati raggiunti con risultati soddisfacenti.

Il piano operativo in uso da parte dell'azienda per mettere in atto il servizio di raccolta porta a porta è stato analizzato e riorganizzato secondo base mensile. La riorganizzazione del piano ha portato ad evidenziare come in determinati mesi dell'anno la forza lavoro necessaria per effettuare il servizio possa servire meno o maggiormente. Questo porta ad evitare il problema di un sovraccarico di forza lavoro impiegata, fornendo un calcolo più preciso e dando la possibilità all'azienda di ridimensionare il servizio stesso garantendo alle utenze una raccolta più efficiente.

Per quanto riguarda i modelli predittivi, si vuole notare come essi siano risultati uno strumento abbastanza affidabile per poter prevedere i valori in esame.

Si devono però considerare come modelli di partenza e quindi non conclusivi, essi sono una discreta base di partenza per l'eventuale sviluppo, eventualmente tramite modelli più evoluti che possono permettere di ottenere delle previsioni molto più affidabili e veritiere di quanto non lo siano già con i modelli ottenuti in precedenza.

Si ricorda, infatti, che lo sviluppo di tali modelli non preclude lo sviluppo di al-

trettanti modelli predittivi per altrettanti determinati parametri differenti da quelli esaminati nella tesi.

Le opportunità future per questo modello sono molteplici e riguardano sia l'ambito operativo che quello decisionale.

Il modello, tramite le stesse operazioni e cambiando i dati in input, può prevedere anche altri parametri. Ad esempio, il modello, in linea di principio, potrebbe prevedere la quantità di rifiuti prodotta in una determinata area e per allocare in modo efficiente le risorse disponibili.

In ottica di sviluppo e miglioramento del modello, sarà importante continuare ad adattarlo alle esigenze del settore e integrandolo con altri sistemi e tecnologie.

L'obiettivo è quello di rendere la raccolta dei rifiuti più efficiente e sostenibile, a beneficio della salute dell'ambiente e del benessere della comunità.

Riferimenti bibliografici

Siti web consultati

- [1] *An Introduction to Statistical Learning with Applications in R*. URL: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf (cit. alle pp. 34, 59–61).
- [2] *Analisi predittiva*. URL: https://www.sas.com/it_it/insights/analytics/predictive-analytics.html (cit. a p. 58).
- [3] *Analisi Serie Storiche*. URL: https://it.wikipedia.org/wiki/Analisi_delle_serie_storiche (cit. alle pp. 32, 53).
- [4] *Analisi serie storiche in R*. URL: <https://cran.r-project.org/doc/contrib/Ricci-ts-italian.pdf>.
- [5] *Bilancio Sostenibilità 2020*. URL: https://www.contarina.it/files/filemanager/source/documenti/Bilancio_2020_stampa.pdf (cit. alle pp. 1, 2).
- [6] *Bilancio Sostenibilità 2021*. URL: https://contarina.it/files/filemanager/source/documenti/Bilancio_sostenibilita_2021%5C%20low.pdf (cit. alle pp. 2, 9).
- [7] *Catasto Nazionale rifiuti ISPRA*. URL: <https://www.catasto-rifiuti.isprambiente.it/index.php?pg=mregione&aa=2020®id=ITALIA> (cit. a p. 7).
- [8] *Catasto Regionale rifiuti ISPRA*. URL: <https://www.catasto-rifiuti.isprambiente.it/index.php?pg=regione> (cit. a p. 7).
- [9] *Catasto Veneto rifiuti ISPRA*. URL: <https://www.catasto-rifiuti.isprambiente.it/index.php?pg=provincia&aa=2020®id=Veneto&idreg=05&p=1> (cit. a p. 7).
- [10] *Contarina SpA*. URL: <https://contarina.it/chi-siamo/contarina> (cit. a p. 1).
- [11] *Cos'è la regressione lineare?* URL: <https://aws.amazon.com/it/what-is/linear-regression/> (cit. a p. 59).
- [12] *Documentazione fitdist()*. URL: <https://www.rdocumentation.org/packages/fitdistrplus/versions/1.1-8/topics/fitdist> (cit. a p. 40).
- [13] *Documentazione funzioni pacchetto stats di R*. URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/> (cit. alle pp. 42, 53, 62).
- [14] *Documentazione R*. URL: <https://www.r-project.org/other-docs.html>.

- [15] *Documentazione readxl()*. URL: <https://cran.r-project.org/web/packages/readxl/readme/README.html> (cit. a p. 35).
- [16] *Indice di correlazione di Pearson*. URL: https://it.wikipedia.org/wiki/Indice_di_correlazione_di_Pearson (cit. alle pp. 42, 44).
- [17] *Metodi di previsione di serie temporali ARIMA*. URL: https://docs.oracle.com/cloud/help/it/pbcs_common/PFUSU/prhist_arima_intro.htm#PFUSU_pp_user_book_217 (cit. a p. 32).
- [18] *Modelli predittivi*. URL: <https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling> (cit. a p. 58).
- [19] *Motivazioni per la scelta di R*. URL: <https://www.northeastern.edu/graduate/blog/r-vs-excel/> (cit. a p. 22).
- [20] *Presentazione aziendale 2019*. URL: <https://www.contarina.it/files/filemanager/source/presentazione-aziendale-2019.pdf> (cit. alle pp. 8, 9).
- [21] *Processo stocastico*. URL: https://www.treccani.it/enciclopedia/processo-stocastico_%5C%28Enciclopedia-della-Matematica%5C%29/ (cit. a p. 34).
- [22] *Rapporto rifiuti urbani Edizione 2021*. URL: https://www.isprambiente.gov.it/files2022/pubblicazioni/rapporti/rapportorifiutiurbani_ed-2021-n-355-conappendice_agg18_01_2022.pdf (cit. alle pp. 1, 2).
- [23] *Regressione Lineare*. URL: <https://www.ibm.com/it-it/analytics/learn/linear-regression#:~:text=La%5C%20regressione%5C%20lineare%5C%20corrisponde%5C%20a,una%5C%20serie%5C%20di%5C%20dati%5C%20accoppiati>. (cit. a p. 58).
- [24] *Regressione lineare in R*. URL: <https://cran.r-project.org/doc/contrib/Ricci-regression-it.pdf>.
- [25] *RStudio*. URL: <https://posit.co/download/rstudio-desktop/#> (cit. a p. 31).
- [26] *Serie Storiche*. URL: https://www.isprambiente.gov.it/files/pubblicazioni/manuali-lineeguida/MLG_84_2013.pdf (cit. a p. 32).
- [27] *Serie storiche, analisi delle*. URL: <http://www.ecostat.unical.it/tarsitano/oldsite/stecm2d1.pdf> (cit. a p. 34).
- [34] *statsmodels 0.14.0 (+734)Stationarity and detrending (ADF/KPSS)*. URL: https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html (cit. a p. 55).

Documenti interni consultati

- [28] Contarina SpA. *Definizione Parametri PO con Banca Dati Associata*. 2019 (cit. alle pp. 10, 13, 14).
- [29] Contarina SpA. *Manuale d'uso LeOCar 2022*. 2022 (cit. a p. 11).
- [30] Contarina SpA. *Presentazione Aziendale 2017*. 2017.
- [31] Contarina SpA. *Presentazione PO38*. 2017 (cit. alle pp. 10, 13, 15).

[32] Contarina SpA. *Presentazione Progetto LeO*. 2020 (cit. a p. [11](#)).

[33] Contarina SpA. *Presentazione Progetto Modello Predittivo*. 2017 (cit. a p. [14](#)).